THE UNIVERSITY
of EDINBURGH

Using Population Data to Understand the Impact of Timing of Birth on Singleton and Twin Pregnancies

Sarah Rose Murray (BSc, MBChB, MRCOG, MSC Epidemiology)

A thesis submitted for the degree of Doctor of Philosophy

School of Clinical Sciences

The University of Edinburgh

2019

Declaration

I declare that this thesis has been written by myself and not submitted for any other degrees. The work submitted is my own work except where work has been submitted as a jointly-authored publication – the work of others has been explicitly indicated in the introductory paragraph of each chapter and appropriate references have been made within the thesis acknowledging the work of others.

Sarah Rose Murray

Acknowledgements

I would like to thank my supervisors, Professor Jane Norman, Professor Jill Pell and Dr Sarah Stock for their support, advice and encouragement throughout my PhD. To my primary supervisor, Professor Jane Norman; for investing in me, setting me up with fruitful collaborations, allowing me to travel, providing sound advice and mentorship and always encouraging and praising me, thank you. My thanks also to my second supervisors, Professor Jill Pell for her invaluable knowledge of epidemiology and Dr Sarah Stock for her expertise in obstetric research. I have been privileged to work with these world-leading experts in obstetric academia and perinatal epidemiology and could not have asked for better role models and supervisors.

I acknowledge and thank the Wellcome Trust for funding my work and the studies in this thesis and for encouraging me to undertake a masters in Epidemiology with the London School of Hygiene and Tropical Medicine. Completing the masters definitely equipped me with the tools necessary to carry out my PhD. I thank my MSC supervisor, Dr Suzanne Penfold, for her input and critique of my MSC dissertation.

I acknowledge Professor Bo Jacobsson, Julius Juodakis and Jonas Bacelis for welcoming me into their lab in Gothenburg, Sweden and for teaching me the intricacies of 'R' as well as laughing at my inability to maintain a tidy workspace. I am very grateful that this was the first study I undertook as part of my PhD as the skills I learned through working in Gothenburg with these statistical experts were extremely useful for the subsequent studies in my PhD.

I thank Dr Sohinee Bhattacharya for welcoming me into the epidemiology unit in Aberdeen and providing me with excellent support and supervision when carrying out the study using the Aberdeen Maternity and Neonatal Databank. I acknowledge Danny MacKay at the University of Glasgow for teaching me about 'GEE', figuring

out my often incomprehensible STATA code and putting up with my many questions about multiple imputation. Thank you to Professor Rebecca Reynolds and Professor Fiona Denison for letting me participate in the Tommy's lab meetings and allowing me to moan about PBPP and the nuances of routine data. I am very grateful to have been part of the Tommy's Centre team and to have such approachable Principal Investigators who were always very happy to offer advice, support and encouragement.

To my family for their support; my parents for their constant belief in me and their unending encouragement, my sisters, friends and work colleagues at the Simpsons for putting up with me through the thesis writing, and to my new husband Andrew for still wanting to marry me through the writing up process and for continually supporting and championing me; thank you.

Sarah Rose Murray
Edinburgh, 2019

Table of Contents

Abstract and Lay Summary

Abstract

Overall in the last two decades there has been a decrease in the average gestational age at delivery. Gestation at delivery is important as it is associated with both short and long-term outcomes for the baby. The gestation the baby is born at affects the risk of perinatal mortality with increased perinatal mortality rates with preterm delivery but also in cases of prolonged pregnancy. In the longer-term, the gestation at delivery can affect the cognitive and school outcomes of the child, especially if born prematurely. Optimising the timing of delivery is therefore an important balance between short and long-term childhood outcomes. This thesis aimed to investigate the impact of timing of delivery on short and long-term outcomes in both singleton and twin pregnancies and the differences between the two using routinely collected maternity data.

In singleton pregnancies, preterm delivery is the largest cause of perinatal and infant mortality with 10% of neonates worldwide born prematurely (<37 weeks). One of the main challenges with preterm delivery is that the aetiology is so wide and largely unknown that implementing the correct interventions for prevention is not yet possible. In this thesis a population cohort study was used to determine the effect of geographical and environmental influences on preterm birth rates in an attempt to identify potential new mechanisms driving preterm birth. In a study of 1,335,802 singleton births, marked differences in the preterm delivery rate were observed across the country with longer gestational ages in urban areas suggesting the effect of urbanity as a potential area for future research. The association of late preterm birth (34-36 weeks) and early term births (37-38 weeks) with long-term cognitive outcomes in the offspring was investigated in the form of a systematic review. In four studies of 35,711 children, infants born at 39-41 weeks had higher cognitive outcome scores than those born at early term (37-38 weeks). This study adds to the growing body of evidence regarding the need to consider both short and long-term outcomes associated with gestation at delivery when planning timing of delivery.

In twin pregnancies the optimum timing of delivery is largely unknown. The short and long-term outcomes according to gestation at delivery were explored initially in a subset of the Scottish population (n=7421) and then in the full Scottish population of 43,133 twins. Short term outcomes investigated included perinatal mortality and long-term outcomes were investigated by record linking the maternity data to the school census data of the child. The optimum gestation for delivery of uncomplicated twin pregnancies is 37 weeks.

To investigate the differences in perinatal mortality between twins and singletons a population cohort study of 2,002,587 infants was performed. Overall twins had a higher rate of stillbirth compared to singletons at all gestational ages from 24 weeks. Neonatal death was higher in twins in the extreme preterm period but lower between 29 and 37 weeks.

In conclusion determining optimum timing of delivery should consider both short and long-term infant outcomes and this information should be used to inform policy makers and when counselling women and families about timing of delivery.

Lay Summary

The purpose of the work presented in this thesis is to find out how gestation (the length of pregnancy) at delivery affects the immediate and future health of babies in singleton and twin pregnancies. The gestation the baby is born at is known to affect the immediate health of the baby with babies born prematurely (less than 37 weeks gestation) at increased risk of death (both stillbirth and infant death). Prolonged pregnancy (greater than 42 weeks) is also associated with an increased risk of death for the baby and is a key reason for inducing labour. In the longer-term, babies born prematurely are known to have higher educational needs at school and lower IQ scores. It is therefore very important to try to determine the optimum gestation at delivery with the greatest balance of short and long-term health complications. In a review of previously published studies conducted as part of this thesis it was found that even babies born at early term (37-38 weeks) had lower IQ scores compared to those born at full term (39-41 weeks) again highlighting the need to balance the short and long-term outcomes when planning timing of delivery.

In singleton pregnancies preterm delivery remains a major cause of infant death with 1 in 10 babies born too soon worldwide. One is the key problems with preterm delivery is that often we do not know the cause of the premature labour and therefore we are unable to prevent it effectively or treat to stop the preterm labour once it has started. Further research is needed to guide new treatment and prevention measures. One of the studies presented in this thesis aimed to determine how the geographical environment a woman was exposed to during pregnancy affected the risk of preterm delivery. The finding that wide geographical variation in preterm birth rates exist and that gestation is longer in urban areas improves our understanding of possible risk factors for preterm birth and will guide future research.

Twins are a high-risk pregnancy requiring specialist obstetric care and they have a threefold increase in death compared to singleton pregnancies. Also, twins are at very high risk of premature delivery with 1 in 2 twins delivering less than 37 weeks

gestation. Although twins only account for 2% of all live births, they account for a large amount of special care baby unit admissions and the rates of twinning are rising worldwide because of the increase in in-vitro fertilisation techniques. Unfortunately, in twin pregnancies there is limited research into the effect of timing of birth on immediate and future health of twins. In this thesis all twin deliveries in Scotland were examined to look at the effect of timing of birth on immediate death of twins and long-term school outcomes of the twins. The lowest risk of both short and long-term problems was found to be in twin babies born at 37 weeks gestation. This information should be used to inform policy makers and clinicians when advising women with a twin pregnancy about the best time for delivery.

In conclusion determining the best time for delivery of both singleton and twin babies should consider both immediate and future health of the children.

Publications and Presentations Relating to this thesis

Poster Presentations

- **Perinatal Outcomes in Twins Compared to Singletons According to Gestation at Delivery: a Population Cohort Study of 2,004,587 Infants in Scotland.**
  Sarah Murray, Suzanne Penfold, Sarah Stock, Jane Norman.
  Royal College of Obstetricians and Gynaecologists World Congress, 2019, London, UK

- **Gestational Age at Delivery of Twins and the Risk of Perinatal Death: a Population Cohort Study.**
  Sarah Murray, Daniel MacKay, Sarah Stock, Jill Pell, Jane Norman.
  Society for Reproductive Investigation Annual Meeting, 2019, Paris, France.

- **Sex Differences in Perinatal Mortality and Morbidity in Twins: a Population Case-control Study in Scotland.**
  Sarah Murray, Suzanne Penfold, Sarah Stock, Jane Norman.
  Royal College of Obstetricians and Gynaecologists Annual Academic Meeting, 2019, London, UK.

- **Long Term Cognitive Outcomes of Early Term (37-38 weeks) and Late Preterm (34-36 weeks) Births: A Systematic Review.**
  Sarah Murray, Susan Shenkin, Kirsten McIntosh, Jane Lim, Benjamen Grove, Jill Pell, Jane Norman, Sarah Stock.
  Society for Reproductive Investigation: Targeting Inflammation, 2018, Edinburgh, UK.

- **Geographical Differences in Preterm Delivery Rates in Sweden: a Population-based Cohort Study.**
  Sarah Murray, Jonas Bacelis, Julius Juodakis, Anna Sand, Jane Norman, Verena Sengpiel, Bo Jacobsson.
  Society for Maternal and Fetal Medicine Annual Meeting, 2017, Las Vegas, USA.

- **Gestational Age at Delivery of Twins and Perinatal Outcomes, a Scottish Cohort Study.**
  Sarah Murray, Sohinee Bhattacharya, Edwin Amalraj Raja, Jane Norman.
  British Maternal and Fetal Medicine Society Annual Meeting, 2017, Amsterdam, The Netherlands.
  Shortlisted for 'Best Poster' prize

- **Using Big Data To Determine Optimal Time of Delivery of Twin Pregnancy.**
  Sarah Murray, Jill Pell, Sarah Stock, Jane Norman.
  The Farr Institute of Health Informatics Research Annual Meeting, 2015, St Andrews, UK.

Oral Presentations

- **Gestational age at delivery of twins and perinatal and childhood outcomes: a population cohort study of 43,133 twins.**
  Sarah Murray, Daniel MacKay, Sarah Stock, Jill Pell, Jane Norman.
  British Maternal and Fetal Medicine Society Annual Meeting, 2019, Edinburgh, UK
  Winner of Best Oral Poster Presentation

- **Gestational Age at Delivery of Twins and the Risk of Perinatal Death: a Scottish Population Cohort Study.**

Sarah Murray, Daniel MacKay, Sarah Stock, Jill Pell, Jane Norman.
Pre-Doctoral Plenary Competition, Academy of Medical Sciences Clinical Academics in Training Annual Conference, 2018, Edinburgh, UK.

- **Perinatal Outcomes in Twins Compared to Singletons According to Gestation at Delivery: a Population Cohort Study of 2,004,587 Infants in Scotland.**
  Sarah Murray, Suzanne Penfold, Sarah Stock, Jane Norman.
  Edinburgh Obstetrical Society, 2018, Edinburgh, UK.

- **Geographical Differences in Preterm Delivery Rates in a Country with a Very High Human Development Index – a Population Study.**
  Sarah Murray, Jonas Bacelis, Julius Juodakis, Anna Sand, Jane Norman, Verena Sengpiel, Bo Jacobsson.
  UK Annual Preterm Birth Conference, 2017, Leeds, UK.

- **Long Term Cognitive Outcomes of Early Term (37-38 weeks) and Late Preterm (34-36 weeks) Births: A Systematic Review.**
  Sarah Murray, Susan Shenkin, Kirsten McIntosh, Jane Lim, Benjamen Grove, Jill Pell, Jane Norman, Sarah Stock.
  UK Annual Preterm Birth Conference, 2017, Leeds, UK.

- **Preterm Birth in Multiple Pregnancy.**
  Sarah Murray, Sarah Stock, Jane Norman
  Preis School Preterm Birth Symposium, 2016, Florence, Italy.

Publications

- Murray SR, Bhattacharya S, Pell JP & Norman JE. Gestational age at delivery of twins and perinatal outcomes: a cohort study in Aberdeen, Scotland. *Wellcome Open Research,* 2019 March, 4:65.

- Murray SR, Juodakis J, Bacelis J, Sand A, Norman JE, Sengpiel V, Jacobsson, B. Geographical Differences in Spontaneous Preterm Birth rates in Sweden: A population-based Cohort study. *ACTA Obstetricia et Gynaecologica Scandinavica.* 2018 August 3; doi: 10.1111/aogs.13455. [Epub ahead of print].

- Murray SR, Stock SJ, Cooper ES, Norman JE. Spontaneous Preterm Birth in Multiple Pregnancy. *The Obstetrician and Gynaecologist*. 2018 January; 20(1): 57-63.

- Murray SR, Stock SJ & Norman JE. Long-term Childhood Outcomes after Interventions for Prevention and Management of Preterm Birth. *Seminars in Perinatology*. 2017 December; 41(8): 519-527.

- Murray SR, Shenkin SD, McIntosh K, Lim J, Grove B, Pell JP, Stock SJ. Long Term Cognitive Outcomes of Early Term (37-38 weeks) and Late Preterm (34-36 weeks) Births: A Systematic Review. *Wellcome Open Research.* 2017 October; 2:101.

- Sarah Murray, Eleanor Whitaker, Jane Norman

  Contribution to book chapter: 'Clinical interventions to prevent preterm birth in the singleton fetus': *Fetal Therapy 2E- Scientific Basis & Critical Appraisal of Clinical Benefits.* April 2019.

- Sarah Murray, Sarah Stock, Jane Norman
  Induction of labour and prolonged pregnancy: StratOG online tutorial March 2016.

Abbreviations

Abbreviation	Full Definition
ACOG	American College of Obstetricians and Gynecologists
ADHD	Attention Deficit/Hyperactivity Disorder
AIC	Akaike's Information Criterion
AMH	Aberdeen Maternity Hospital
AMND	Aberdeen Maternal and Neonatal Databank
ART	Assisted Reproduction Technology
ASQ	Ages and Stages Questionnaire
BMI	Body Mass Index
BSID	Bayley Scores of Infant Development
CI	Confidence Interval
CMAQ	Community Multiscale Air Quality Model
CRH	Corticotrophin Releasing Hormone
CSV	Comma Separated Values
eDRIS	Electronic Data and Research Innovation Service
FDR	False Discovery Rate
GEE	Generalised Estimating Equations
ICD	International Classification of Disease
IOL	Induction of Labour
IPD	Individual Patient Data
IQ	Intelligence Quotient
ISD	Information Services Division
IUGR	Intrauterine Growth Restriction
MAR	Missing at Random
MBRRACE	Mothers and Babies: Reducing Risk Through Audits and Confidential Enquiries across the UK
MCAR	Missing Completely at Random
MNAR	Missing Not At Random
MRI	Magnetic Resonance Image
NHS	National Health System
NICE	National Institute for Care Excellence
NICU	Neonatal Intensive Care Unit
NND	Neonatal Death
NNU	Neonatal Unit
NRS	National Registry of Statistics
NSS	National Services Scotland
OR	Odds Ratio
PBPP	Public Benefit and Privacy Panel for Health and Social Care
PMR	Perinatal Mortality Rate
PPROM	Preterm Premature Rupture of Membranes
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Premature Rupture of Membranes

PROSPERO	The International Prospective Register of Systematic Reviews
PTD	Preterm Delivery
QIC	Quasi-likelihood Independence Model Criterion
RCT	Randomised Controlled Trial
REC	Regional Ethics Committee
RFM	Reduced Fetal Movements
ROBANS	Risk of Bias for Nonrandomized Studies
RR	Relative Risk
RSE	Robust Standard Errors
SCBU	Special Care Baby Unit
ScotXed	Scottish Exchange Educational Data
SEN	Special Educational Need
SMBR	Swedish Medical Birth Registry
SMFM	Society Of Maternal and Fetal Medicine
SMR02	Scottish Morbidity Record 02
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SSBID	Scottish Stillbirth and Infant Death Survey
WASI	Wechsler Abbreviated Scale of Intelligence
WISC	Wechsler Intelligence Scale for Children
WHO	World Health Organisation

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction 1: Gestation at Delivery of Twin and Singleton Pregnancies

1.1 Gestation

Gestation at delivery refers to the length of pregnancy that has been completed when a baby is born. A pregnant woman is defined as being 'at term' when her pregnancy reaches a duration of 37 weeks. The 'term' period however covers a wide range of gestations from 37-42 weeks and in 5-10% of women their pregnancy will extend beyond 42 weeks gestation (Olesen, Basso and Olsen 2003). Preterm delivery is defined by the World Health Organisation (WHO) as babies born alive before 37 weeks of gestation are completed. Globally preterm delivery occurs in 10% of neonates and the rates are rising worldwide (Blencowe et al. 2012). Preterm delivery in the Scotland occurs in around 6.6% of singleton pregnancies and up to 65% of twin pregnancies and the percentage of babies that are born preterm has increased in recent years from (from 6.1% in 2008/9 in singletons and 55% of twins in 2008/9)(Scotland 2018).

Gestation at delivery is important to consider because as the pregnancy continues beyond term the risk of the baby dying in the uterus or soon after delivery increases (Hilder, Costeloe and Thilaganathan 1998) in both twin and singleton pregnancies. Routine ultrasound scanning early in pregnancy to determine gestational age via a measurement of crown-rump length has therefore been recommended for all pregnancies (twin and singleton pregnancies) in national policies (NICE 2010). Accurate identification of gestational age has been shown to reduce the need for induction of labour (IOL) in post-term pregnancy (Whitworth, Bricker and Mullan 2015).

Preterm delivery is also important to consider because prematurity is the most common single cause globally of perinatal and childhood mortality (Ferrero et al. 2016a). For infants who survive preterm delivery there is an increased risk of neurological disability with the risk increasing with decreasing gestational age.

1.2 Optimum Gestation at Delivery

1.2.1 Short-Term Offspring Outcomes in Singleton Pregnancies

National policy in the UK recommends offering IOL to women with singleton pregnancies at 41 weeks gestation if birth has not occurred in order to reduce the short term offspring outcome of perinatal mortality (NICE 2010). IOL is a common procedure in the UK occurring in up to 28% of pregnancies with prolonged pregnancy being one of the most common indications (Blotkamp et al. 2018). The UK national policy is based on a large Cochrane systematic review and meta-analysis of 17 trials (7,407 women) which demonstrated a significant reduction in perinatal mortality in the IOL groups (at 40 weeks) compared to expectant management groups (relative risk [RR] 0.31, 95% confidence intervals [95% CI] 0.12-0.88)(Gülmezoglu et al. 2012). IOL was also associated with a reduction in rates of caesarean section (RR 0.69, 95% CI 0.81-0.97) and a reduction in meconium aspiration syndrome (RR 0.50, 95% CI 0.34-0.73)(Gülmezoglu et al. 2012, Wood, Cooper and Ross 2014).

There are a number of groups of women in whom earlier IOL has been recommended in order to reduce the short-term offspring outcome of perinatal mortality. These groups include women with pre-existing diabetes (Boulvain, Stan and Irion 2001), women with prelabour rupture of membranes (PROM)(Carroll 2010) and women with hypertensive disorders of pregnancy (Visintin et al. 2010).

There is also an increasing body of evidence to support earlier IOL with advanced maternal age (defined as greater than 35 years of age) however in contrast to randomised controlled trial (RCT) data available for prolonged pregnancy, the data

investigating the effect of advanced maternal age is observational and hence bias cannot be eliminated. An American observational study (>5 million women) demonstrated that advanced maternal age was an independent risk factor for antenatal and intrapartum stillbirth (rate of stillbirth at 41 weeks in women <35 years 0.75 per 1000 compared to 2.5 per 1000 in women >35 years)(Reddy, Ko and Willinger 2006). A recent RCT was undertaken to address the question of earlier IOL in older women with a primary outcome of caesarean section rate and a secondary outcome of adverse neonatal outcome (Walker et al. 2016). The trial of 619 women found that IOL among women of advanced maternal age had no increased risk of caesarean section compared to expectant management (RR 0.99, 95% CI 0.87-1.14) and no increased risk of adverse neonatal outcomes (admission to neonatal unit [NNU] 0.88, 95% CI 0.36-3.06). Following this RCT, an observational cohort study was performed to look at perinatal mortality as the primary outcome in women over 35 years old who had been induced compared to those expectantly managed. This study of 77,327 women found that IOL at 40 weeks was associated with a lower risk of perinatal mortality (adjusted [adj.] RR 0.33, 95% CI 0.13-0.80) and meconium aspiration syndrome (adj. RR 0.52, 95% CI 0.35-0.78) compared to those women who were expectantly managed (Hannah et al. 2017). The study concluded that routinely offering IOL at 40 weeks in women over the age of 35 may result in lower overall rates of perinatal mortality. This evidence is particularly pertinent to the UK population where the proportion of pregnancies in women over 35 years has risen substantially from 8% in 1985 to 20% in 2000 (Dhanjal and Kenyon 2013).

There is also emerging evidence that earlier IOL in all women (not just those listed above as being high risk) may be associated with lower rates of short-term adverse outcomes in offspring. A large observational study of 1,271,549 women compared outcomes of IOL at each week of gestation from 37 weeks to 41 weeks compared to expectant management. The study concluded that the IOL group was associated with a reduction in perinatal mortality compared with the expectant management group at each week of gestation (at 40 weeks gestation adj. OR 0.39, 95% CI 0.24-0.63)(Stock et al. 2012). This study, however, did show an increase in NNU admission in the IOL

group (at 40 weeks adj. OR 1.14, 95% CI 1.09-1.20). Subsequently, a further RCT has recently been published investigating the outcomes of IOL in low risk nulliparous women from 39 weeks onwards. In this RCT 3,062 women were assigned between 38 and 38+6 weeks gestation to undergo IOL at 39 weeks and compared to 3,044 women who were expectantly managed. IOL in the former group of women did not result in a decrease in the composite adverse perinatal outcome (RR 0.80, 95% CI 0.64-1.00)(Grobman et al. 2018) but did result in a decreased risk of caesarean section (RR 0.84, 95% CI 0.76-0.93). A recently published meta-analysis of 6 cohort studies of 66,019 women undergoing elective IOL and 584,390 undergoing expectant management upheld the main RCT findings (Grobman and Caughey 2019). In this review the IOL group had significantly lower rates of caesarean section (RR 0.83, 95% CI 0.74-0.93), peripartum infection (RR 0.53, 95% CI 0.39-0.72), meconium aspiration syndrome in the neonate (RR 0.49, 95% CI 0.26-0.92), respiratory morbidity in the neonate (RR 0.71, 95% CI 0.59-0.85) and perinatal mortality (RR 0.27, 95% CI 0.09-0.76) compared to the expectantly managed group.

Overall, in recent years, a number of studies have shown a decrease in the mean gestational age at delivery (mean gestational age decreased from 40 weeks in 1994 to 39 weeks in 2004)(Gyamfi-Bannerman 2011) and a mean of 0.4 weeks in a population cohort study of 43,217 singletons (Gibson et al. 2015). It is not entirely clear why this occurred but could be due to the emerging body of evidence surrounding earlier IOL to reduce perinatal mortality as mentioned above. The number of IOLs for the indication of reduced fetal movement (RFM) at term has also increased, in recent years, due to the association between RFM and stillbirth (Stacey et al. 2011). However, the recently published stepped-wedge cluster RCT assessing the introduction of a RFM package of care with the aim of reducing perinatal mortality did not result in a reduction in stillbirth rates in the intervention group (adj. OR 0.90, 95% CI 0.75-1.07)(Norman et al. 2018). Hence the evidence for IOL for RFM alone remains unclear.

1.2.2 Long-Term Offspring Outcomes in Singleton Pregnancies

Although there is evidence supporting IOL at gestations earlier than 41 weeks in singleton pregnancies to reduce adverse short-term perinatal outcomes, it is also important to consider the effects of timing of delivery on long-term outcomes in the offspring of which there is a limited evidence. Preterm infants have increased rates of neurological compromise compared to term born infants (Moore et al. 2012). However, as established above the period of term refers to a broad range of gestational weeks and early term (37-38 weeks) births are increasingly common (Gyamfi-Bannerman 2011). Gestation at delivery has a strong dose-dependent relationship with the risk of having special educational need (SEN) at school with a progressive decrease in SEN requirement with increasing gestational age (MacKay et al. 2010). In this large observational study of 407,503 schoolchildren the risk of SEN was lowest at 41 weeks. A strategy of early IOL to reduce perinatal morality may reduce the risk of perinatal death (which is a rare outcome) but result in an increase in developmental compromise in the child.

1.2.3 Short-Term Offspring Outcomes in Twin Pregnancy

Although twin pregnancies account for only 3% of live births they are associated with significantly higher adverse outcomes than singleton pregnancies. Twin pregnancies have a threefold increase in perinatal mortality compared to singletons (Manktelow et al. 2014) and have a preterm delivery rate of approximately 50% compared to 5% in singletons (ISD Scotland 2009). Twin pregnancy is associated with significantly higher risk for both the mother and the babies. Maternal consequences of twin pregnancy include increased rates of pre-eclampsia, pregnancy induced hypertension, gestational diabetes and anaemia (Duckitt and Harrington 2005, Chamberlain 1991). Fetal consequences of twin pregnancy include increased rates of preterm birth, intrauterine growth restriction, congenital anomalies and stillbirth (Visintin et al. 2011, Boyle et al. 2013). Monochorionic twins [twins who share a placenta] carry the additional risk of twin-twin transfusion syndrome which occurs in approximately

15% of monochorionic pregnancies and carries a significant risk of fetal morbidity and mortality (Kilby 2017). Monochorionic twins therefore have much more extensive monitoring compared to dichorionic pregnancies and so chorionicity is an important factor to consider when looking at perinatal outcomes because the rates of adverse outcomes are much higher in monochorionic twins compared to dichorionic twins.

Despite accounting for only 3% of live births, twin infants account for around 15% of special care baby unit and NNU admissions (Harrison and Goodman 2015). Twin pregnancies also require more frequent monitoring and contact with healthcare professionals in view of the excess obstetric risks listed above. This results in an increased economic burden on the NHS and the costs involved in the care of a twin pregnancy are estimated to be three times the costs involved in a singleton pregnancy(RCOG 2017). Despite the UK policy of single embryo transfer to reduce the rates of multiple pregnancy (Fields et al. 2013); in 2011 the twin pregnancy rate was still increasing with up to 24% of successful in vitro fertilisation (IVF) procedures at that time resulting in a multiple pregnancy (Visintin et al. 2011). In contrast to singletons conceived through IVF procedures, twins conceived via IVF procedures do not appear to have increased rates of adverse perinatal outcome compared to naturally conceived twins (Helmerhorst et al. 2004).

Optimising the timing of delivery is a key strategy in minimising perinatal death in both twins and singletons. National policy in the UK recommends elective birth from 37 weeks in dichorionic twin pregnancy (two placentae and two separate chorions [external fetal membranes]) and elective birth from 36 weeks in monochorionic twin pregnancy (shared placenta with separate chorion [monochorionic diamniotic] or shared chorion [monochorionic monoamniotic]). The national policy is based on evidence from two large population studies of fetal death in multiple pregnancy demonstrating an increased risk of perinatal mortality after 37 weeks gestation (Kahn et al. 2003, Minakami and Sato 1996). Subsequently, a systematic review of RCTs

and observational studies (29,685 dichorionic and 5,486 monochorionic twins) has upheld the main conclusions of the population studies.

1.2.4 Long-Term Offspring Outcomes in Twin Pregnancy

In contrast to the limited data emerging for singleton pregnancy, there is a paucity of evidence regarding the optimum gestation of delivery for twins in terms of long-term offspring outcomes. Most studies on long-term outcomes of twins compare twins to singletons as opposed to comparing twins born in certain gestational age categories (Maria et al. 2013, Babatunde et al. 2018, Tsou et al. 2008) and therefore the effect of gestation at delivery on long-term developmental outcomes in twins is largely unknown.

1.3 Preterm Delivery

1.3.1 Short-Term Offspring Outcomes in Singleton Pregnancy

Preterm delivery is the most common cause of perinatal and infant mortality worldwide. Globally, 10% of neonates are born prematurely (<37 weeks)(Blencowe et al. 2012). Infants who are born preterm and survive are subsequently at risk of long-term neurological disability (Chang et al. 2013). The economic burden of preterm birth is therefore substantial given that so many babies are affected worldwide.

The aetiology of preterm birth is wide and implementing the correct intervention is challenging as it is unrealistic that any single intervention will work across all aetiologies. Over 8 different mechanisms of preterm birth have been identified, as shown in Figure 1-1. Indeed, a recent study of 4,100,000 births from five countries concluded that known risk factors only account for approximately one third of preterm births thus highlighting the need for further research into the aetiology of spontaneous preterm birth in singletons (Ferrero et al. 2016a).

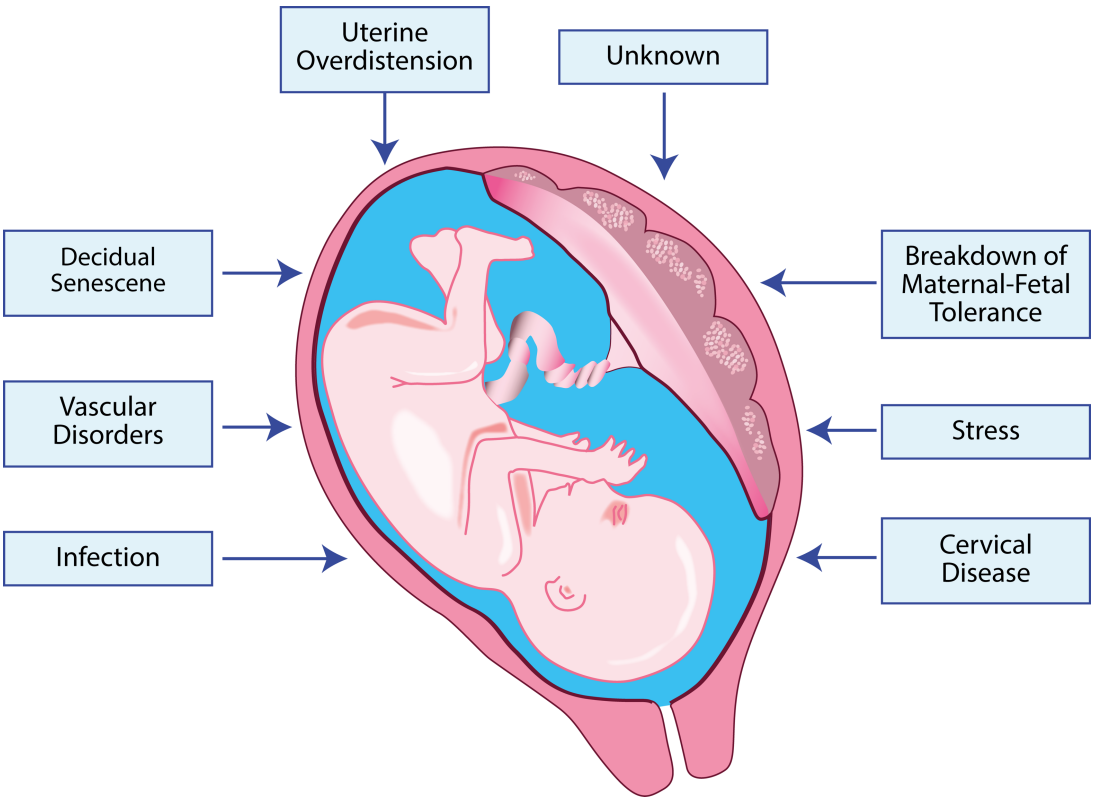


Figure 1-1: Proposed mechanisms of disease implicated in spontaneous preterm birth adapted from Romero, Dey and Fisher 2014.

There are three main strategies that have been developed to try to prevent preterm birth. These interventions include cervical cerclage (a purse-string suture to strengthen and tighten the cervix inserted between 12 and 24 weeks and removed at 37 weeks), vaginal and intramuscular progesterone (started between 16 and 22 weeks, the vaginal preparation is the only one available in the UK and is usually prescribed as a once daily pessary with the aim of maintaining uterine quiescence by preventing functional withdrawal of progesterone) and cervical pessaries (a silicon ring sitting around the cervix to support and tilt it posteriorly inserted at 18-22 weeks and removed at 37 weeks)(Stock and Ismail 2016). These interventions are shown in Figure 1-2. A large study of 39 countries published in 2013 investigated the potential

reduction in preterm birth using the available interventions. The study concluded that implementation of known interventions would produce a relative reduction in preterm birth of only 5% (Chang et al. 2013). The best intervention for preterm birth prevention therefore remains unclear (Stock and Ismail 2016). Further research into the aetiology of preterm birth and the available interventions to prevent preterm birth is urgently needed.

1.3.2 Short-Term Offspring Outcomes in Twin Pregnancy

As stated above twin pregnancy is associated with a threefold greater perinatal mortality rate compared to singletons. Much of this increased morbidity is thought to be driven by prematurity. The aetiology of preterm birth in twins is likely to be multifactorial and, in many cases, different to singletons. However, in a similar way to singletons the multifactorial aetiology means that it is also very difficult to treat and prevent preterm birth in twins. Proposed pathophysiology processes involved in preterm birth in twins include intrauterine infection, cervical insufficiency and increased uterine stretch/distension. In a multiple pregnancy, due to the increased placental mass, there is increased secretion of mediators such as corticotrophin-releasing hormone (CRH) and surfactant protein-A, both of which are known to stimulate myometrial contractility and may contribute to the high preterm birth rates in twins (Stock and Norman 2010). As well as increased rates of spontaneous preterm labour twin pregnancy is also associated with increased rates of iatrogenic preterm delivery with approximately one third of all preterm multiple pregnancy deliveries being medically indicated (Fuchs and Senat 2016). The proposed mechanisms of preterm birth in multiple pregnancy are shown in Figure 1-2.

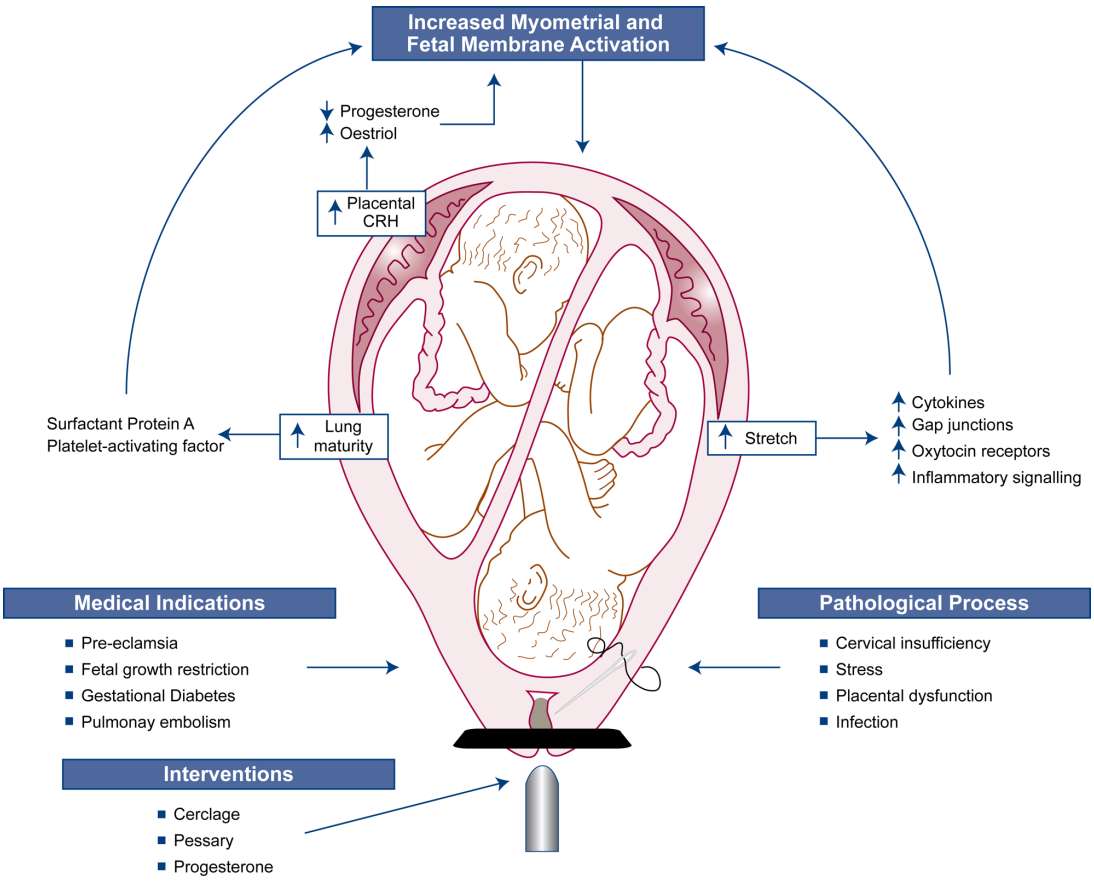


Figure 1-2: Potential mechanisms of preterm birth in multiple pregnancy and potential interventions adapted from Stock and Norman 2010.

Interventions available for preterm birth prevention in multiple pregnancy are similar to the interventions available for singletons but with differing success. Cervical cerclage is associated with an increased risk of preterm birth in multiple pregnancies (RR 2.15, 95%CI 1.15-4.01)(Fuchs and Senat 2016) and is therefore not routinely recommended. There is also no clear evidence of benefit for the use of progesterone for preterm birth prevention in multiple pregnancy (Schuit et al. 2015, Norman et al. 2009) and the cervical pessary is currently only used in multiple pregnancy in the research setting (Stock and Ismail 2016).

1.4 Differences in Perinatal Mortality between Twins and Singletons

Twins have a greater risk of perinatal mortality compared to singletons (Manktelow et al. 2014) and a higher preterm birth rate (68% in twins compared to 6.5% in singletons)(ISD Scotland 2018). Because of the increased risks of these and other adverse outcomes in twin pregnancy, twins are monitored more closely than singleton pregnancies and receive a large amount of obstetric input (Kilby 2017). Previous studies have shown that, despite a greater risk of perinatal mortality overall in twins compared to singletons, the perinatal mortality in preterm twins is actually lower than that of singletons (Minakami and Sato 1996, Vasak et al. 2017). The lower preterm perinatal mortality rate in twins compared to singletons is thought to be biologically plausible as it is hypothesised that the aetiology of preterm labour varies between the two pregnancy types. In twin pregnancy the most common aetiology for preterm birth is thought to be uterine stretch (Figure 1-2) compared to infection in preterm singleton pregnancies (Figure 1-1). If this is the case, twin babies born preterm may therefore be born "in better condition" than those singleton preterm babies born in the context of infection or pre-eclampsia. Babies born in the context of infection may, themselves, be infected at delivery and become unwell more quickly compared to babies born in the context of no infection (e.g if the cause of preterm delivery was uterine stretch). Likewise, if a baby is born in the context of preterm pre-eclampsia it is often growth restricted at delivery and again more unwell than a baby born prematurely because of uterine stretch (although pre-eclampsia is also more common in twins than in singletons). This has potentially important implications for how we manage antenatal care for singletons at risk of preterm birth.

Chapter 2

Introduction 2: Long-term Outcomes of Preterm Delivery in Singleton and Twin Infants

The following materials have been published in Seminars in Perinatology in 2017 (Murray, Stock and Norman 2017) under the title 'Long-term childhood outcomes after interventions for prevention and management of preterm birth' by Dr Sarah R Murray (SM), Dr Sarah J Stock (SS) and Professor Jane E Norman (JN). SM prepared the first draft of the manuscript under the guidance of JN. SS provided the section on antenatal corticosteroids (not included below). All authors provided critical insight for the manuscript and approved the final version

In summary, this review aimed to identify and synthesise the literature on long-term outcomes of preterm delivery in both singleton and twin pregnancies. Babies born preterm are known to have increased rates of neurological disability in later life and this increases with decreasing gestational age. This trend appears to continue across the gestational weeks of term and as discussed in **Chapter 1** even late preterm births (34-36 weeks) are associated with increased risk of cerebral palsy (RR 3.1, 95% CI 2.3-4.2)(Teune et al. 2011) compared to term deliveries. The aim of interventions to reduce preterm birth in both singleton and twin pregnancies is to prolong pregnancy and this is presumed to improve the health of the babies (Stock and Ismail 2016). These interventions for preterm birth prevention form a major focus of obstetric practice however it is still uncertain whether delaying delivery results in improved health outcomes in the children. Therefore, even if the interventions for preterm birth achieve the intermediate outcome of increased gestational age at delivery, less is known about whether they achieve long–term health benefits.

2.1 Long-Term Outcomes of Progesterone for Preterm Birth Prevention

Progesterone is available as an intramuscular injection of 17 α- hydroxylase caproate (only licensed in the USA) or a vaginal progesterone preparation (the only available progesterone product in the UK, but not licensed either in USA or Europe for preterm birth prevention). It is currently recommended for use for preterm birth prevention in the UK NICE guideline for certain specific categories of women (singleton pregnancies at high risk of preterm birth). Biological plausibility for the use of progesterone comes from the concept that uterine quiescence is maintained throughout pregnancy and labour is thought to occur as a result of a functional withdrawal of progesterone (Norwitz, Robinson and Challis 1999). Some work has been done in recent years regarding the safety of the use of progesterone for the prevention of preterm birth (O'brien 2012, O'brien 2015). The most recent review published in 2016 by O'Brien and Lewis (O'Brien and Lewis 2016) of the safety of 17 α- hydroxylase caproate concludes that its use is contraindicated in multiple pregnancies because of the risk of adverse immediate neonatal events (RR 1.21, 95% CI 1.03 – 1.43 for a composite outcomes of death and severe morbidity) and that in singletons further research is needed to determine its safety. Studies investigating the effectiveness of progesterone in preventing preterm birth have conflicting results. A recently published Cochrane review and individual patient data (IPD) meta-analysis of the use of progesterone in singleton pregnancies demonstrated it was an effective agent in preventing preterm birth in women with previous preterm birth and a short cervix (Romero et al. 2012, Dodd et al. 2013). The recently published OPPTIMUM trial, the largest RCT (n = 1228) to date of vaginal progesterone versus placebo for prevention of preterm birth demonstrated no difference in gestational age at delivery between the two groups (Norman et al. 2016). Nevertheless, despite this controversy, progesterone is used widely throughout the world for preterm birth prevention and therefore information about the long-term childhood neurological outcomes is crucial for counselling women about its use.

The OPPTIMUM trial reported on childhood outcomes at age two (n = 869) using the Bayley Score of Infant Development (BSID). There were no statistically significant differences in the scores between the progesterone and placebo group reported with

a difference in means of -0.48 (95% Confidence intervals [CI] -2.77 to 1.81). Analysis of secondary outcomes showed (non statistically significant) higher rates of death from trial entry to age of two in the progesterone group (3% compared with 4%, OR 1.28 [95% CI 0.66, 2.51], p = 0.48) and a (non statistically significant) higher incidence of moderate to severe neurodevelopment disability (9% compared with 12%, OR 1.48 [95% CI 0.98, 2.33], p = 0.087). A study by Northen *et al.*(Northen et al. 2007) performed the longest follow up study to be done in singletons with a mean age at follow-up of 48 months (n = 270). This was a follow up of the National Institute of Child Health and Human Development Maternal-Fetal Networks Study of 17 α- hydroxylase caproate as part of a multicentre placebo-controlled trial (Meis et al. 2003). The initial study demonstrated a significant reduction in the rate of spontaneous preterm birth but the follow-up study reported that, despite 17 α-hydroxylase caproate apparently preventing preterm birth, scores of the 'Ages and Stages' questionnaire (ASQ) did not differ significantly between the progesterone and the placebo groups being within normal ranges in both (ASQ score below cut-off on at least one area 27.5% in the progesterone group compared with 28% in the placebo group, p = 0.92).

In multiple pregnancies a placebo-controlled RCT of vaginal progesterone published by Rode *et al.*(Rode et al. 2011), the PREDICT trial, reported on long-term infant follow-up. The infants were assessed by ASQ at six and 18 months after the expected date of delivery (n = 1,050). There were no statistically significant differences found in the mean scores between the progesterone group and the placebo group (ASQ mean score at six months 215 compared with 218, p = 0.45 and mean ASQ score at 18 months 193 compared with 194, p = 0.89). The STOPPIT (Norman et al. 2009) RCT also compared vaginal progesterone with placebo in twin pregnancies and a follow-up study published in 2015 investigated the effect of vaginal progesterone on childhood outcome (McNamara et al. 2015). The mean age at follow-up was 55.5 months and the 'Child Development Inventory' was used to measure childhood outcome (n = 759). There was no evidence of difference between the progesterone-exposed and the placebo-exposed twins (Child development Inventory score below

cut-off on at least one area 30% compared with 35%, p = 0.66), equally there was no difference in the overall health index of the groups (Health Utilities Index rating 'excellent' 88% compared with 90%, p = 0.51). A further follow up of the PREDICT babies has recently been published by Vedel *et al.*(Vedel et al. 2016) providing the longest follow-up to date of children aged eight years (n = 989). The primary outcomes investigated by this study were neurophysiological development of the children assessed by the ASQ and admissions and diagnoses up to eight years of age using medical records of the children. The study did not report any harmful effect of exposure to progesterone in terms of diagnoses and admissions (n = 989). A statistically significantly higher mean ASQ score in the progesterone group compared to the placebo group was reported (mean total score 269, [standard deviation SD 28.2] compared with 261.7 [SD 31.4], p = 0.03) but of note the scores were only received on 437 of the children (45.8% response rate but no differences found on maternal characteristics of responders and non-responders).

As well as a putative effect on delaying the onset of labour, it has been proposed that progesterone may have a direct beneficial effect on the fetal brain. Progesterone has recently been investigated because of its potential therapeutic use in acute traumatic brain injury in adults. Biologically beneficial effects are thought to be feasible because progesterone is widely distributed throughout the central nervous system with some neuroprotective properties having been demonstrated (Singh and Su 2013). However, despite initial positive studies, a Cochrane review published in 2016 which included five studies of 2,392 participants, did not find any evidence that progesterone was superior to placebo in reducing death or disability in adults with traumatic brain injury (Ma et al. 2016). Similar to the follow up studies of exposure in utero, there were no reports of adverse effects of progesterone seen. Just as progesterone is thought to be present in abundance in the central nervous system of adults, it is also thought to be present in the fetal brain during pregnancy and after birth. It has been hypothesized that prophylactic progesterone could have beneficial effects in fetuses at high risk of brain injury such as those with a low estimated fetal weight or intrauterine growth restriction (Vedel et al. 2016). However further

research is necessary, as a 2016 study published in 2016 by Willing *et al.* found that in the rodent model in utero exposure to 17 α- hydroxylase caproate caused detrimental effects in the fetal brain resulting in impaired cognitive flexibility in adult life (Willing and Wagner 2016).

In summary the follow-up of the effect of fetal exposure to progesterone in utero ranges from six months to 8 years of age and does not show any evidence of harm, although the amount of evidence is limited. Clinicians and pregnant women at high risk of preterm birth should make individual decisions on whether the best available evidence suggests that progesterone is appropriate for them.

2.2 Long-Term Outcomes of the Cervical Pessary for Preterm Birth Prevention

In Eastern European countries cervical pessaries have been used for many years for preterm birth prevention (Arabin and Alfirevic 2013). However, in the UK and the USA, cervical pessaries for preterm birth prevention are only recommended for use in a research setting (Arabin and Alfirevic 2013, Stock and Ismail 2016), reflecting uncertainty in their benefit. Three studies have shown a reduction in preterm birth in women with a short cervix in both singletons and/or twins (Goya et al. 2012, Goya et al. 2016, Liem et al. 2013a) but two studies showed no benefit of the pessary in preventing preterm birth (Nicolaides et al. 2016). A systematic review of both RCTs and cohort studies published in 2013 showed potential benefit of the pessary in both twins and singletons but highlighted the need for more research (Liem et al. 2013b) due to the ongoing conflict as to the effectiveness of the pessary. A systematic review and meta-analysis of cervical pessaries in twin pregnancies with a short cervix showed no reduction in spontaneous preterm birth rates or adverse neonatal morbidity (Saccone et al. 2015). Many further RCTs are ongoing/planned (>20 listed as of December 2016, clincialTrials.gov) and therefore it is important to consider the long-term effects of this treatment modality.

The only long-term follow-up study of the cervical pessary was performed at three years of age in children born in the ProTwin RCT of women with a multiple pregnancy and a short cervix (van 't Hooft et al. 2018). The results are currently only available as an abstract but are due to be published soon alongside the 4-year follow-up. This follow-up study used the BSID-III scores of infant development and also looked at deceased and disabled children at three years of corrected age (n = 171). A higher survival without disability was found in the pessary group versus controls (92.4 vs 73.8%, p = 0.006) and among survivors there were no statistically significant differences in the scores of children having been exposed to the pessary compared to standard care. The study concluded that use of the cervical pessary for preterm birth prevention does not appear to be associated with adverse neurological outcomes for children.

2.3 Long-Term Outcomes of Cervical Cerclage for Preterm Birth Prevention

Cervical cerclage is one of the oldest surgical techniques described for preterm birth prevention (RCOG 2015). Recent studies looking at trends in the use of cervical cerclage in the United States have shown an overall decline in its use (Suhag et al. 2015) however it is currently recommended for use in UK and USA guidelines for preterm birth prevention (Sarri et al. 2015, Ressel 2004). The evidence surrounding the effectiveness of cervical cerclage has been synthesized in a Cochrane review published in 2012 of 12 RCTs (3328 women) and reported an overall risk reduction 0.80 (95% CI 0.69 – 0.95) in preterm birth with the use of cervical cerclage (Alfirevic et al. 2012). Despite there being good evidence for the use of cerclage in singletons for preterm birth prevention it does not appear to have the same effect in multiple pregnancies and in some studies has a trend towards harm (Rafael, Berghella and Alfirevic 2014, Saccone et al. 2015). There is a paucity of evidence reporting long-term infant neurodevelopmental outcomes of cervical cerclage for preterm birth prevention despite the longevity of its use. No studies were identified assessing outcomes beyond the neonatal period in the two Cochrane reviews that evaluated the

use of cervical cerclage for preventing preterm birth in both singletons and in twins (Rafael et al. 2014, Alfirevic et al. 2012) highlighting a need for further research in this area.

2.4 Long-Term Outcomes of Tocolytics for Preterm Birth Prevention

Tocolytics are used to prevent spontaneous preterm birth in women with symptoms of preterm labour who require in-utero transfer or who have not completed a course of antenatal corticosteroids. In a systematic review of 17 trials (Gyetvai et al. 1999) tocolytics (beta-agonists, indomethacin atosiban and ethanol) were associated with prolonged pregnancy for 24, 48 hrs and seven days but without any benefit on neonatal morbidity and mortality. In a systematic review and network meta-analysis of tocolytic agents, nifedipine and atosiban were found to have similar efficacy and side-effects but costs of nifedipine were lower (Haas et al. 2012).

Romero *et al.* followed up the children to the age of one in an RCT of tocolysis. An increased risk of infant death was reported in the atosiban group (RR 6.15, 95% CI 1.39 – 27.22) however these results should be interpreted with caution as the sample size was small (as demonstrated by the wide confidence intervals) and there was randomization bias as more women with very preterm labour (<26 weeks) were given atosiban (Romero et al. 2000).

The group of studies entitled ‘alleviation of pregnancy outcome by suspending tocolysis in early labour’ (APOSTEL) trials have been conducted in the Netherlands to look at the outcomes of different tocolytics and methods of preterm birth prevention. A follow up study of the APOSTEL-II trial (nifedipine versus placebo for maintenance tocolysis) performed the ‘Ages and stages’ questionnaire in 170 infants in the trial at two years of age. Infants of mothers who had received nifedipine maintenance therapy had a higher incidence of fine motor problems (22.2% versus 7.6%, OR 3.42 [95% CI 1.29-9.14], p = 0.01) but a lower incidence of poor problem solving (21.1% versus 29.1%, OR .27 [95% CI 0.08, 0.95], p = 0.04). The study concluded that there was no clear evidence to support the use of nifedipine for

maintenance tocolysis as it showed no benefit in the immediate outcomes (Roos et al. 2013) or in the two year follow-up study of the infants (van Vliet et al. 2016). The APOSTEL III RCT of nifedipine versus atosiban for threatened preterm birth has recently been published (van Vliet et al. 2016). Overall the paper reported similar perinatal outcomes in both groups and did not perform long-term follow-up of the infants however there was a trend towards an increased risk of perinatal death in the nifedipine group (5% compared with 2%, RR 2.20 [95% CI 0.91-5.33]). Further studies are planned comparing atosiban with placebo and the long-term follow up is required to provide more information on the outcomes of tocolytics for preterm birth prevention.

2.5 Long-Term Outcomes of the Use of Antibiotics for Preterm Birth Prevention

A key hypothesis in the aetiology of preterm birth is ascending infection and intrauterine infection. The use of antibiotics may therefore be useful in treating inflammation in an attempt to reduce the chance of intrauterine infection.

A Cochrane review of 22 trials involving 6,800 women with preterm prelabour rupture of membranes (PPROM) showed a reduction in immediate outcomes of perinatal infection and chorioamnionitis but there was no reduction in perinatal mortality demonstrated (Kenyon, Boulvain and Neilson 2010). Co-amoxiclav (amoxicillin-clavulanic acid) use was associated with an increased risk of necrotizing enterocolitis and therefore the recommendations list erythromycin as the antibiotic of choice. A follow-up study of the ORACLE I trial which compared the use of erythromycin and/or co-amoxiclav with placebo for women with PPROM (Kenyon, Taylor and Tarnow-Mordi 2001) was published in 2008 and included in the Cochrane review (Kenyon et al. 2008). This follow-up study consisted of a health questionnaire and recorded national curriculum test results for children at age seven years with a primary outcome of any level of functional impairment using the Multi Attribute Health Status classification (n = 3,298). There was no difference in the proportion of

children with functional impairment or medical conditions in the antibiotic-receiving group compared to no antibiotic-receiving group (38.3% compared with 40.4%, OR 0.91, 95% CI 0.79-1.05). In terms of school outcomes both groups performed equally and the study concluded that the prescription of antibiotics did not have a clinically or statistically significant effect on health of the children at seven years. Although current literature supports the use of antibiotics in women with PPROM it is important to counsel woman of the lack of long-term effect of its use in this category. Given the negative consequences of antibiotic administration in women with intact fetal membranes (see below), accurate diagnosis of PPROM is important.

There is evidence of harm associated with the use of antibiotics for preterm birth prevention in women with intact membranes without overt signs of infection. The most recent Cochrane review published in 2013 included data from 14 studies of 7,838 women (Flenady et al. 1998). A reduction in maternal infection was reported but there was no difference found in perinatal death between the two groups or in rates of preterm birth. Neonatal deaths were higher in the antibiotic group (RR 1.57, 95% CI 1.03 – 20.4). The long-term outcomes reported in the systematic review were dominated by results of the ORACLE II follow-up study (Kenyon et al. 2008). Similar to the ORACLE I follow up study, children were assessed at age seven for functional impairment, health status and educational outcomes (n = 3,196). Overall there was an increased risk of functional impairment in the antibiotic group (OR 1.18, 95% CI 1.02 – 1.37). There was also an increased risk of cerebral palsy in the antibiotic group (erythromycin OR 1.93, 95% CI 1.21 – 3.09, co-amoxiclav OR 1.69, 95% CI 1.07 – 2.67). The reason for the increased risk of behavioural impairment and cerebral palsy associated with antibiotic use is not clear. One hypothesis is that the increased risk may be a direct effect of the antibiotic exposure. Another relates to keeping the fetus in a hostile environment, prolonging the pregnancy and leading to fetal brain injury (Kenyon, Hagberg and Norman 2013). Neither of these pathways has been proven and further research is necessary.

Azithromycin has also been trialled for the prevention of preterm birth. Azithromycin has broad-spectrum antibacterial properties and is effective against *ureaplasma* species that are commonly found in association with preterm birth. One large trial randomised 2,297 women in Malawi to placebo or azithromycin for preterm birth prevention. No significant differences in preterm birth rates were reported between the azithromycin and placebo group (van den Broek et al. 2009). There were no long-term effects published with this trial and the authors combined the primary outcome of preterm birth <37 weeks in a meta-analysis with 7 other trials and found no benefit of a reduction in preterm labour (RR 1.02, 95% CI 0.86 – 1.22).

In summary the evidence regarding long-term outcomes of antibiotics for preterm birth prevention is dominated by the ORACLE follow-up studies. In women with PPROM although there is evidence of a benefit in immediate outcomes of maternal and fetal infection there are no long-term benefits associated with their use. In women with intact membranes there is evidence harm in the long-term effects of their use with an increased risk of functional impairment and cerebral palsy. The link between the unique vaginal microbiome and preterm delivery continues to be an active area of research (Aagaard et al. 2012, Romero et al. 2014b). As demonstrated by the ORACLE studies, it is imperative that these studies include long-term follow up of the children.

In summary although various strategies of preterm birth prevention exist there is a paucity of evidence available regarding the long-term outcomes of these strategies as demonstrated in Figure 2-1.

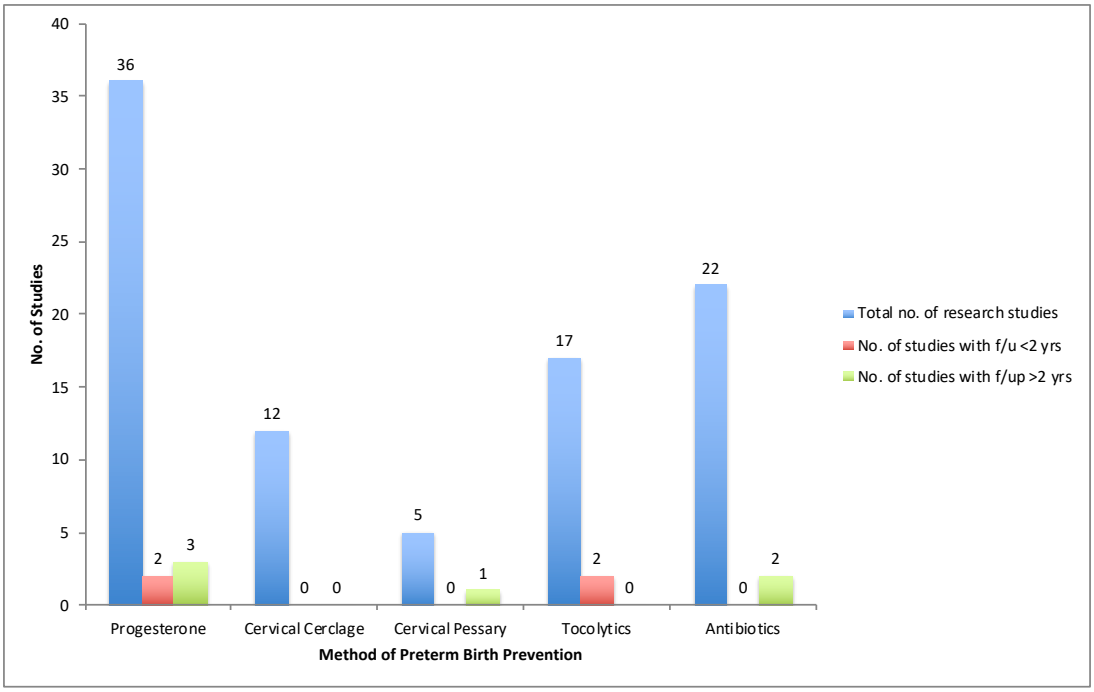


Figure 2-1: Proportion of studies investigating preterm birth prevention with long-term follow up

2.6 Chapter Conclusion

2.6.1 Further Relevant Research Since Manuscript Publication

Progesterone for Preterm Birth Prevention

The PROGRESS study (vaginal progesterone pessaries for pregnant women with a previous preterm birth to prevent neonatal respiratory distress syndrome) was published by Crowther *et al*. in 2017 (Crowther et al. 2017). This trial of 740 women with a previous preterm delivery, pregnant with either twin or singleton pregnancies, demonstrated that the risk of respiratory distress syndrome was similar in the progesterone and the placebo groups (RR 0.98, 95% CI 0.64-1.49). The PROLONG (hydroxyprogesterone caproate to reduce preterm birth) trial of 1,707 women with a previous spontaneous singleton preterm birth has also been completed in 2019 and did not demonstrate any statistically significant difference in the neonatal morbidity

and mortality composite index between the treatment and placebo arms (5.4% versus 5.3%, p = 0.84)(Blackwell et al. 2018). These studies add to the existing body of evidence suggesting that the role of progesterone is likely to be limited (Norman and Bennett 2017).

In contrast to the two large studies above demonstrating no benefit of progesterone, Jarde *et al.* published a network meta-analysis comparing progesterone, cerclage and the cervical pessary for preterm birth prevention in singleton pregnancies (Jarde et al. 2017). The meta-analysis of 36 trials of 9,425 women concluded that progesterone was the best intervention for preventing preterm birth in singleton pregnancies at risk of preterm birth and neonatal mortality and other significant morbidity sequalae.

In summary, the evidence for the use of progesterone for preterm birth prevention is conflicting. The EPPPIC (Evaluating Progestogens for prevention of preterm birth international collaboration) IPD meta-analysis has been initiated to address the uncertainties in the use of progesterone for preterm birth prevention (Stewart et al. 2017). The EPPPIC study will amalgamate data from RCTs of progesterone versus placebo in women with singleton and multiple pregnancies and various risk factors for preterm birth with a primary outcome of preterm birth or fetal death.

Cervical Pessary for Preterm Birth Prevention

Two further RCTs have been published demonstrating a reduction in spontaneous preterm birth with the cervical pessary. Saccone et al.(Saccone et al. 2017) performed a RCT of 300 women with a singleton pregnancy and a short cervical length and found a reduction in preterm birth rates below 34 weeks gestation (7.3% vs. 15.3%, RR 0.48, 95% CI 0.24-0.95). Following this trial a systematic review and meta-analysis was published showing an overall reduction in preterm birth with the cervical pessary in singleton pregnancies (RR 0.46, 95% CI 1.82-2.31)(Pérez-López et al. 2019). A recent RCT of 132 women with a singleton pregnancy at high risk of preterm birth (short cervix) randomised to cervical pessary or standard care was published by Merced et al. in 2019 and has upheld the result of the meta-analysis

(Merced et al. 2019). This study found a reduction in the preterm birth rates in the pessary, compared to the control, group (RR 0.51, 95% CI 0.27-0.97).

The long-term follow-up of the cervical pessary use in twin pregnancies has now been published by van't Hooft *et al*.(van 't Hooft et al. 2018) demonstrating a reduction in the cumulative incidence of death or survival with a neurodevelopmental disability in the pessary group compared to the control group (OR 0.26, 95% CI 0.09-0.73). Amongst the children who survived, there was no difference in cognitive, language or motor development. The study concluded that when used in women with a twin pregnancy and a cervical length of less than 38mm, use of the cervical pessary strongly improved survival of the children without affecting their neurodevelopment at three years.

Chapter 3

Introduction 3: The use of Population Data to Study Pregnancy Outcomes

3.1 Population Sample

Population data refers to datasets that are too large or too complex to be analysed by simple methods of data analysis (De Mauro, Greco and Grimaldi 2016). In healthcare, this can refer to the use of mandatory routinely collected data on populations of thousands or millions. Routine data comes from ongoing data collections systems in health and administrative social services. The main benefit of using routine data is that it is readily available largescale comprehensive and non-selective population data which broadly speaking does not require additional sampling techniques and which is already available retrospectively since the year the collection system commenced. By reducing the need for sampling, the target population is actually the population studied and hence there is a low risk of selection bias in the participants. There are some caveats to this population approach in pregnancy, for example if a woman has a home birth then she does not generate a hospital record as she has not been admitted to hospital and is therefore not captured in the routine data, this event is rare however and estimated to be <1% of all births in Scotland but needs to be considered when carrying out population data studies. As well as reducing the risk of selection bias, population samples are usually large enough to give us the power to study rare outcomes such as perinatal mortality in rare populations, for example twins. As twins are only 3% of all livebirths it would require a long period of recruitment to generate the large sample required in an RCT setting to be able to study rare outcomes such as perinatal mortality. Indeed the 'Twin Birth Study' carried out to compare the risk of fetal or neonatal death or serious morbidity from planned caesarean section compared to planned vaginal delivery took 8 years to recruit the required sample size of 2804 women at a high cost (Barrett et al. 2013). By using routinely collected population data there is also a lower risk of recall or

information bias as the data is collected remotely from any research question. Population data studies are therefore large, efficient, low cost and generally low risk to undertake.

3.2 Record Linkage

Another benefit of using routinely collected healthcare data is the ability to perform record linkage to other health records and cross-sectoral records for research purposes. Thus, providing an efficient way to perform follow-up cohort studies of long-term offspring outcomes without the need to recruit and retain patients and gain individual patient consent. In Scotland data linkage of maternity records to school education records has been carried out successfully to study pregnancy exposures and long-term school outcomes (Wood et al. 2013, MacKay et al. 2010). As well as linking to education data, successful linkage has been performed to prescription data and employment records to allow long-term follow up based on the exposure of a particular medication, for example the relationship between exposure to medication for attention-deficit/hyperactivity disorder (ADHD) and education and health outcomes (Fleming et al. 2017). Using population data also allows aspects of fetal programming to be investigated through data linkage techniques. Datasets in Scotland, such as the Aberdeen Maternity and Neonatal Databank (AMND) have been accurately collecting maternity data since 1950 and it is now possible to look at in-utero exposures on long-term offspring health and indeed death. For example, a cohort study performed using routinely collected maternity data from the AMND record linked to the offspring health records, found maternal obesity to be associated with an increased risk of premature death in adult offspring (Reynolds et al. 2013).

3.3 Limitations of Population Data

There are limitations to the use of population data to perform research and many of these are covered in the methods and data chapters throughout this Thesis. Key limitations include:

- Missing data – this refers to data that is not collected but also data that is collected but not complete. Missing entries within covariates can lead to a loss in study power if omitted from the multivariable analysis, bias if the missing is not missing at random or residual confounding if the decision is made to omit the variable from the multivariable analyses because of the amount of missing data. As the data for routinely collected studies is not prospectively collected there are inevitable potential confounders that are not recorded and therefore unable to be addressed in the multivariable analyses. One such variable is chorionicity in the twin studies using SMR02 which can then be a source of residual confounding.
- Data quality – This refers to both the completeness of the data (as discussed above) but also the accuracy of the data. The overall quality of the data provided routinely can be a limitation if it has not been formally assessed for errors. Errors in data entry can lead to measurement error which is a form of information bias. Misclassification (differential and non-differential) of either the exposure or outcome will again result in a form of information bias. Most of the databases used throughout this Thesis have been subject to regular quality control assessment.
- Unrecorded temporal changes occur in routine datasets – these changes include the fact that medical practice and obstetric care will have changed over time in many of the datasets and these changes are not formally recorded. For example, in the Aberdeen Maternity and Neonatal Databank used in this **Chapter 6** the maternity data has been collected over a long period of time, since 1950. In that time, it is likely that both obstetric and neonatal care has both changed and improved leading to improvements in perinatal death rates. However, there are also changes that would lead to increases in perinatal death in twins as there has been an increase in the rate of twinning due to IVF procedures and this in turn will lead to an increase in preterm deliveries. These changes in care and the incidence of twinning over time need to be considered in the multivariable analyses.

- Difficulties with case ascertainment – similar to temporal changes in care leading to differences over time, there can be changes in the definitions of cases in routinely collected data over time. This can result in changes in the number of reported cases which can be artefactual if it is indeed just a change in the diagnostic criteria or a change in a laboratory test resulting in a change in the case definition. For example, the gestational age of viability has changed over time having previously been 28 weeks it is now 24 weeks meaning there may have been more miscarriages recorded in earlier years that would now be classed as live births.
- Difficulties with data linkage – although data linkage to other administrative datasets is one of the key benefits of using routinely collected data it can also be very difficult to link data on individuals in the different datasets. Unique identifiers are needed to be able to successfully link the data and this then leads to potential issues with confidentiality as the individuals need to be identified to be successfully linked. Twins pose a potentially larger problem for data linkage as they have the same postcode, date of birth and mother id which are usually the unique identifiers used to perform individual data linkage. If the twins are non-identical then fetal sex can be used for perform the record linkage.

3.4 Epidemiological Study of Perinatal Mortality

One of the fundamental aims of epidemiological research is to measure and study events in the population at risk, analytical epidemiology then focuses on the factors associated with the outcomes. To do this we need to obtain numerator data (number of new cases of the event/disease) and denominator data (population at risk of the event or disease. Perinatal mortality rate is defined by the World Health Organisation (WHO) as the number of late fetal deaths (28 weeks gestation or more) added to the number of early neonatal deaths (deaths within the first 7 days of life) in the same year divided by the number of late fetal deaths added to the number of life births in the same year. This is generally used as an overall summary statistic of the risk of

perinatal death. Early studies used perinatal mortality rate at each gestational week as the estimate of probability of perinatal death at term (Smith 2001). There are some problems with this definition for estimating perinatal risk or probability however, as it often does not relate to the population at risk. A number of different methods have therefore been described for estimating the risk of perinatal death at different gestational age weeks. For example, the risk of stillbirth at 42 weeks gestation only affects pregnancies that reach 42 weeks and the denominator is all ongoing pregnancies rather than just babies born in that week of gestation. It is therefore now widely accepted that the risk of stillbirth at a given week should be a ratio of the number of stillbirths to the total number of ongoing pregnancies at the start of the week as that is the group at risk of the event (Yudkin, Wood and Redman 1987). Some studies have been performed in the past estimating perinatal risk based on time to event methods with gestation as the time factor and stilbirth the event (Hosmer Jr and Lemeshow 1999), the benefit of using time to event analyses is that censoring (births not resulting in stillbirth) can be accounted for in the analysis. Other studies have estimated the risks of stillbirth cumulatively using the 'cumulative incidence method' with live birth as the competing event as it eliminates the possibility of stillbirth (Naimi and Auger 2016).

For neonatal complications, different methods of determining risk also exist. The population at risk/denominator is different from stillbirths and generally accepted to be the number of pregnancies that are delivered in a given week (Smith 2005). Competing risk analyses of expectant management versus elective delivery have been previously performed and define the risk of perinatal death at a given gestational week as the difference between the risk of stillbirth and the risk of neonatal death for deliveries in that week (Cheong-See et al. 2016).

Stratification by gestational age is key when studying perinatal mortality because the major cause of neonatal death is prematurity (Smith 2005). It is more important to stratify by gestational age or investigate it as a covariate in the regression models rather than simply including gestational age as a covariate. This is because the effect

of gestation is not constant over the continuum of pregnancy with much higher rates of perinatal death at lower gestational age weeks compared to later gestations (Smith 2005).

3.5 Epidemiological Study of Twin Pregnancy

Twin pregnancy is more complex to study with epidemiological techniques than singleton pregnancy as the twin offspring are not independent. The sample size required to study rare outcomes in twin pregnancies such as perinatal death are large and given that twins account for only 3% of the population would take a long time to recruit the necessary sample. As mentioned previously, one of the key benefits of using big data to study twins is the population-based approach resulting in a large unselected twin population. Due to the challenging analyses of twin pregnancy it is in fact often part of the exclusion criteria in many epidemiological studies. Observational studies of twin pregnancy that have been performed in the past have used statistical tests for unpaired data, therefore assuming independence of the twin infants (Minakami and Sato 1996, Sairam, Costeloe and Thilaganathan 2002). However, the outcomes from twin infants are not independent as the twins are genetically more similar (especially if identical twins) and experience the same in-utero exposures as each other making them more alike than infants from different pregnancies (Carlin et al. 2005). Assuming independence will lead to error in the estimated effect with confidence intervals that are too narrow to account for the paired outcomes (Gates and Brocklehurst 2004). Epidemiological studies involving twin pregnancies should therefore account for the clustering/paired effect of the twin infants using cluster trial designs in the case of intervention studies or using statistical techniques to perform a paired or correlated analysis in epidemiological studies (conditional multivariate logistic regression [or McNemar's test for univariate analyses], logistic/cox regression with robust standard errors, random effects modelling or generalised estimating equation multivariable regression analysis). Another issue with twin observational studies performed in the past is the lack of adjustment for birth order of the twins. Second twins are known to be at increased

risk of perinatal death and delivery-related complications and therefore failing to include this potential confounder may lead to residual confounding or error in the estimated effect (Smith, Pell and Dobbie 2002). Another issue specifically relating to twin pregnancies with regard to studying perinatal death is the timing of the antepartum death. In singleton pregnancies if the pregnancy ends in an antepartum fetal death, standard management is to medically expedite birth by inducing labour and the birth will most often occur within the same week and this will be counted as a stillbirth in that week of gestation (Smith 2005). In twins, however, if there is an antepartum fetal death of one twin before 37 weeks the pregnancy is often allowed to continue to benefit the second twin until around 37-38 weeks gestation. If the stillbirth is then recorded as occurring at 37-38 weeks when in fact it occurred much earlier in pregnancy this will lead to an overestimation of the number of stillbirths at 37-38 weeks gestation (Smith 2005). All of these factors must be considered and attempted to be addressed when conducting epidemiological studies in twins.

3.6 Aims, Hypothesis and Outline of Thesis

Mean gestational age at delivery is decreasing and both short-term and long-term offspring outcomes are important aspects to consider when determining optimum timing of birth: studies investigating both short and long-term offspring outcomes are therefore an important focus for current research, especially in twins where the long-term evidence is sparse.

Preterm birth is the leading cause of perinatal and infant mortality Worldwide. However, there still remain many unanswered questions regarding the differing aetiologies of preterm birth. This, therefore, makes it challenging to guide and develop new interventions for prevention and management of preterm birth. When the baby is born preterm it is at risk of long-term neurological problems and increased risk of having SEN at school. Further studies are required to investigate the different aetiologies of preterm birth to try to reduce the number of babies born early and therefore reduce the long-term neurological burden.

Due to the large sample sizes required to study rare events, such as perinatal mortality, RCTs are not feasible and population-based observational studies using routinely collected data are the mainstay of investigation.

This Thesis aimed to investigate the impact of timing of delivery on short and long-term outcomes in both singleton and twin pregnancies and the differences between the two by using routinely collected maternity data. For the yet unanswered question of the optimum timing of delivery of twins, the overarching hypothesis was that similar to singletons, twins would have a dose response effect with SEN according to gestation at delivery and that the optimum timing would be a balance between short-term outcomes of perinatal death and long-term education outcomes. This thesis contains an overall methods chapter, describing methods used in any part of the thesis (**Chapter 4**), and five results chapters (**Chapters 5 - 9):**

In **Chapter 5**, the associations between maternal geographical location and the environment on singleton preterm birth rates are explored as aetiological factors or potential new mechanisms to explain the differences in preterm birth rates across the country of Sweden. It was hypothesised that Sweden being a country with a relatively homogenous population and comprehensive healthcare system should not have large geographical differences in preterm birth rates across the country. A number of urban versus rural environmental and socioeconomic factors were then investigated to try to explore the geographical differences in more detail.

In **Chapter 6**, the relationship between gestation at delivery and the short-term offspring outcome of perinatal death in twins is explored using a retrospective cohort study design. The analysis was stratified by chorionicity, an important risk factor for perinatal death. In a subsequent exploratory analysis, the association between IVF conception status and perinatal death is also assessed. It was hypothesised that the week of gestation associated with the lowest risk of perinatal death would be earlier

than that of singletons and the perinatal death rates would be higher in monochorionic twins compared to dichorionic twins.

In **Chapter 7**, using a population study of routinely collected Scottish maternity data linked to childhood educational data the optimal week of gestation for birth of twins in terms of both short and long-term offspring outcomes was determined. It was hypothesised that this date would be earlier in twins compared to singletons, but that SEN would have the same dose-response effect with gestational age in twins as it has in singletons.

In **Chapter 8**, a population cohort study was used to determine the differences in short-term perinatal outcomes of stillbirth and neonatal death between twins and singletons according to gestational age at delivery. It was hypothesised that in line with previous research, twins would have higher perinatal mortality overall but that preterm twins would actually have lower rates of perinatal mortality compared to preterm singletons.

In **Chapter 9**, a systematic review looking at long-term cognitive outcomes of late preterm (34-36 weeks) and early term (37-38 weeks) singleton deliveries is presented. It was hypothesised that in comparison to children born at full term (39-41 weeks) children born late preterm (34-26 weeks) and early term (37-38 weeks) would have lower long-term cognitive outcome scores.

Chapter 4

Materials and Methods

This chapter describes the data sources and the rationale behind the statistical methods used to generate the results throughout this PhD.

4.1 Data Sources

As stated in the introduction the benefits of using routinely collected data to study pregnancy outcomes are now well recognised. The datasets used throughout the PhD are described below. Record linkage was used for the study described in **Chapter 7** to provide a follow up cohort of twin children; the techniques involved in this process are also described below.

4.1.1. Swedish Medical Birth Registry

The Swedish Medical Birth Registry (SMBR, **Chapter 5**) has electronically collected data prospectively from the first antenatal visit in each pregnancy in Sweden since 1973. The register is held and maintained by the National Board for Health and Welfare, Sweden. It is mandatory for all health care providers to report to the register. The SMBR contains information on gestational age, demographic factors including age, sex, deprivation category and ethnicity as well as lifestyle and medical history including smoking status at booking, obstetric history, obstetric complications, birthweight of the baby and information about the peripartum period. The National Population Register is used to validate the SMBR which contains data on >99% of all births in Sweden (approximately 100 000 births per year). The SMBR is subject to regular quality control exercises on an annual basis and a quality analysis of the register has been previously published (Cnattingius et al. 1990).

4.1.2 Aberdeen Maternity and Neonatal Databank

The Aberdeen Maternity and Neonatal Databank (AMND, **Chapter 6**) was established in 1950 by Professor Sir Dugald Baird to study the physiology and pathology of pregnancy. The data is held and controlled by the University of Aberdeen Department of Obstetrics and Gynaecology. The AMND holds data from all the Aberdeen Maternity Hospital births from 1949 to the present day. Information recorded includes maternal demographics, obstetrics history, obstetric complications, delivery details and information about the baby including birthweight, sex and Apgar scores. The AMND has a very stable population in a defined geographical area with low rates of migration The Aberdeen Maternity Hospital is the only facility in the Aberdeen area and >99% of residents deliver in the hospital. Initially the data was collected on purpose-written punched cards and the AMND was converted to an electronic database in 1986. The AMND contains data on over 200 000 pregnancies. The database is subject to regular quality assurance exercises and completeness checked against the National Health Service (NHS) records. Information on zygosity for twins has been available since 1968. A full data profile of the AMND has been published recently (Ayorinde et al. 2016).

4.1.3. Scottish Maternity Data

Scottish Morbidity Record 02

The Scottish Morbidity Record 02 (SMR02, **Chapters 7-8**) is a database of all inpatients and day cases discharged from obstetric services across NHS Scotland and it is mandatory for healthcare providers to complete. There are approximately 125 000 records generated per year of which around 50% are deliveries. Information is collected on maternal demographics, obstetric history and complications, delivery information and neonatal information including birthweight, sex and Apgar score. The data is held by the NHS National Services Scotland (NSS), Information Services Division (ISD). The SMR02 is subject to regular quality assurance and has been more than 99% accurate for hard coded data since the 1980s(Cole 1980).

Scottish Stillbirth and Infant Death Survey

The Scottish Stillbirth and Infant Death Survey (SSBID, **Chapters 7-8**) collects information on all stillbirths and infant deaths registered with the National Records of Scotland (NRS). Registration is mandated by law. After receiving the death registration data from NRS the SSBID collects additional information including post-mortem reports and relevant case summaries of stillbirths neonatal deaths and late fetal deaths. This database is held by NHS NSS, Information Services Division (ISD). As the database is formed from NRS records the information is complete and the number of unregistered stillbirths and infant deaths likely to be extremely small.

4.1.5 Scottish Exchange of Educational Data

The Scottish Exchange of Educational Data (ScotXed, **Chapter 7**) is part of the Education Analytical Services Division within the Learning and Justice Directorate of the Scottish Government. The Data is therefore held by the Scottish Government. The ScotXed database contains information on the pupil school census which is conducted annually by all local authority-run primary, secondary and special schools. Information is collected on attendance, exclusions, school leavers destinations (higher education and training, employment and unemployment) and whether the child has a record of SEN and the reason for this (defined by the Scottish Department of Education as having additional support needs arising from four main factors: disability or health, learning environment, family circumstances and social and emotional factors). In **Chapter 7,** SEN was defined as a learning difficulty that requires special educational provision; physical/motor or sensory impairment, intellectual difficulties, autistic spectrum disorders or emotional/behavioural problems in line with previous studies (MacKay et al. 2010). The ScotXed also holds the SQA examination database which maintains a record of all children who have been entered for a qualification and the results obtained. The SQA database is a recognised and quality controlled basis for comparing different qualifications of the

children including standard, intermediate and higher grades. It is mandatory for local authorities to complete the school census.

4.1.3 Data Extraction and Record Linkage

The SMBR data for **Chapter 5** was extracted by the National board for Health and Welfare, Sweden.

The AMND data for **Chapter 6** was extracted by the data co-ordinators of the AMND.

For **Chapters 7** and **8** the SMR02 and SSBID maternity data was extracted by the Electronic Data and Research Innovation Service (eDRIS) part of ISD Scotland and the data placed in the National Services Scotland (NSS) National Safe Haven. The data in the safe haven was anonymised for the analysis. A user agreement was signed before access to the safe haven was granted for all users and analysis was completed on remove servers at the University of Edinburgh.

Data Linkage

For **Chapter 7,** data linkage processes were required in order to record link the Scottish maternity data to the ScotXed education data. The maternity health records were linked to the ScotXed school census by probability matching techniques based on the Howard Newcombe techniques of record linkage (Newcombe and Kennedy 1962). The NRS indexing team used the personal identifiers provided by ISD Scotland and matched them to the population spine using complex algorithms. ISD used the secure transfer protocol 'Globalscape' to pass the NHS numbers for the cohorts to the NRS indexing team. The NRS indexing service use a secure transfer protocol 'Thru' to receive the personal identifiers from the Scottish Government ScotXed databases. ISD were then provided with the file back with a unique record

ID number specific to that dataset created by the NRS indexing team. The personal identifiers were removed and 'Globalscape' was used to place the datasets in the National Health Service (NHS) NSS National Safe Haven for the analyses. Data Sharing Agreements and Privacy Impact Assessments were completed by both parties (Scottish Government and the researchers at the University of Edinburgh) prior to data linkage (**Appendix 1**)

Figure 4-1 illustrates the flow of the data through the linkage process performed by ISD, NRS indexing team and the Scottish Government ScotXed Department.

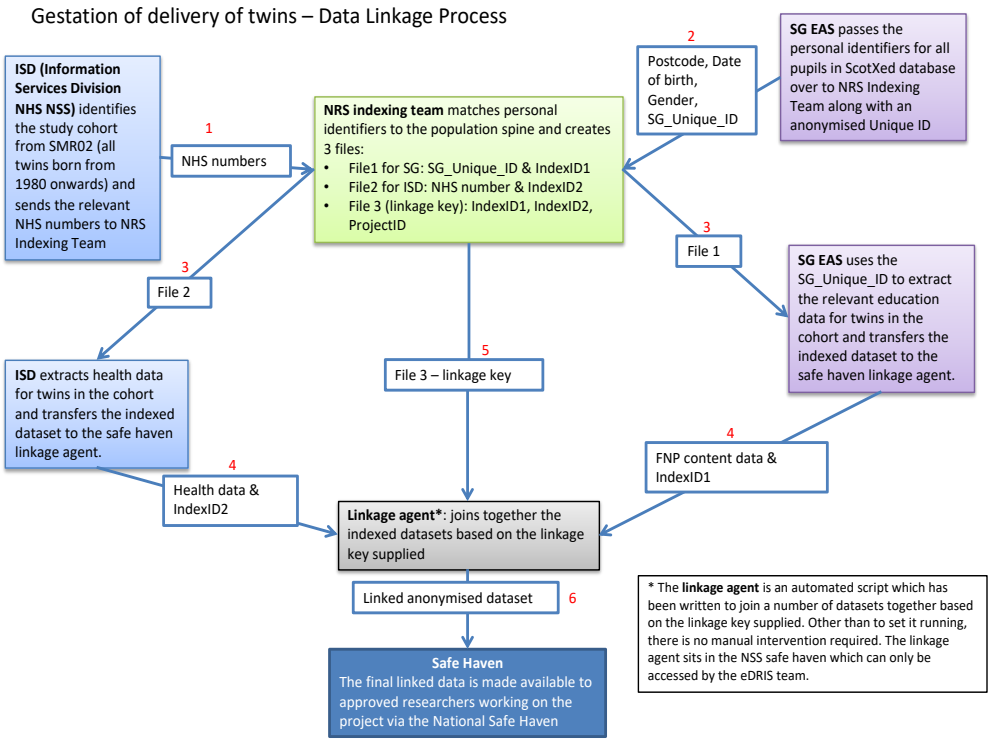


Figure 4-1: Flow Chart of the Data Linkage Process produced by eDRIS for the data sharing agreement, appendix 1 (1-6 detailing the order of steps involved in the linkage process)

4.2 Ethics

4.2.1 Ethics and Other Approvals

The work presented in the chapters of this thesis relate to the following studies and approvals:

1. Geographical Differences in Preterm birth rates in Sweden: A population-based Cohort Study (**Chapter 5**). The study was approved by the Regional Ethical Review Board in Gothenburg, Sweden (968-14).

2. Gestation of delivery of twins– influences on perinatal mortality and morbidity and childhood educational outcomes (**Chapter 6**). The study was approved by the AMND steering committee (approved protocol 104490/Z/14/Z and approval letter in **Appendix 2**).

3. Gestation of delivery of twins – influence on perinatal mortality and morbidity and childhood educational outcomes (**Chapter 7**). The study was approved by the NHS Scotland Public Benefit and Privacy Panel for Health and Social Care (PBPP ref 1516-0252, approval letter in **Appendix 2**) and the NHS South-East Scotland Multi-Centre Research Ethics Committee (NHS REC ref 15/SS/0197, approval letter in **Appendix 2**) and the Scottish Government analytical services division (Data Sharing Agreement in **Appendix 1**).

4. Preterm perinatal mortality in twins compared to singletons: a population study (**Chapter 8**). The study was approved by the NHS Scotland Public Benefit and Privacy Panel for Health and Social Care (PBPP ref 20171117:1718-0132, approval letter in **Appendix 2**) and the NHS South-East Scotland Multi-Centre Research Ethics Committee (NHS REC ref 15/SS/0197, approval letter in **Appendix 2**) and the London School of Hygiene and Tropical Medicine MSc Research Ethics Committee (LSHTM ref 14968, approval letter in **Appendix 2**)

4.2.2 Data Protection and Confidentiality

Access to the data in **Chapters 7** and **8** was through the NSS National safe haven where access is logged and monitored for unusual activity. The safe haven is a virtual private network which provided a Scotland wide research platform for analysis of electronic patient data. An eDRIS user agreement was signed by the investigating team (**Appendix 2**) prior to accessing the patient data. The data was all anonymised prior to being placed in the National safe haven. eDRIS have a Statistical Disclosure Control Protocol detailing how files are released from the safe haven. Files extracted had no counts with less than 10 individuals and files were reviewed by two members of the eDRIS staff before being released.

4.3 Data Preparation

4.3.1 Raw Data Preparation

The data for all the analyses (**Chapters 5-8**) were provided in comma separated values (CSV) delimited files. For all the maternity datasets, the data was provided in 'wide' format with each row corresponding to a separate delivery and each delivery had their own anonymised study ID number. Using this study ID number the SMR02 and the SSBID files could be merged.

For the education records, the ScotXed files were extracted and provided in 'long' format with multiple rows per child because the school census and the examinations database were collected on an annual basis. The same unique index was used in the ScotXed and the maternity databases to allow merging of the two files.

The CSV files were imported into STATA MP, version 14.1 (stata corporation) for the analyses in **Chapters 6-8** and R (version 3.4.1) for **Chapter 5**.

4.3.2 Data Cleaning

The first step in the analyses in **Chapters 5-8** was data cleaning. Data cleaning is the process of displaying and inspecting the data to identify obvious outliers (unusual values of a variable) or possible errors in the data (Kirkwood and Sterne 2010). Data cleaning is particularly important for large routinely collected datasets which by nature are subject to misclassification due to inaccurate recording of variables either by human error or at the data entry stage. Systematic misclassification of exposure, outcome or potential confounding variables can then lead to bias in the results of the analyses. The first step in the cleaning of large datasets is to examine the distribution of each of the exposure, outcome and potential confounding variables to check for errors. Potential confounders were defined as variables associated with both the exposure and the outcome but not on the causal pathway (Kirkwood and Sterne 2010). For categorical variables, this involves checking that the categories make sense whilst for numerical/continuous variables it involves checking that the values do not lie out-with an expected range (for example a maternal height of 1600 cm).

Once outliers have been identified the two following solutions were untaken – the whole record was removed, or the nonsense value of the variable was replaced by a missing value code. The decision to remove records or replace with missing was made depending on the clinical importance of the variable, for example, gestation was the main exposure variable in **Chapters 6-8** and the main outcome variable in **Chapter 5** therefore if this was recorded as an outlier (for example a gestational age at delivery of 100 weeks) or if it was recorded as missing then the whole record was removed. Largely the same process and criteria for the maternity data outlier identification was adopted for all **Chapters 5-8** and is summarised in **Table 4-1**.

For the education data used in **Chapter 7,** data cleaning consisted of identifying the pupils at risk of having a record of special educational need (SEN) therefore removing the maternity records where the child was born before the school census was collected

or was born too early to yet have a school record available (<4 or >19 years at the time of the school census). Children who were born out with Scotland but went to Scottish schools were also excluded as they had no maternity records to link the education data to.

Table 4-1: The method of outlier identification for core variables

VARIABLE	OUTLIER DEFINITION	COMMENT/RATIONALE
GESTATION AT DELIVERY	>43 weeks records excluded Missing records excluded	Main exposure variable in Chapters 6-8 Outcome variable in Chapter 5
MATERNAL AGE	<13 and >60 years of age excluded Missing records excluded	Potential confounder
MATERNAL HEIGHT	<100 and >200 converted to missing values due to the large amount of missing values	Potential confounder
PARITY	>14 excluded Missing records excluded	Potential confounder
BIRTHWEIGHT	<400g or >5000g excluded Missing records excluded	Potential Confounder
OUTCOME OF PREGNANCY	Missing records excluded	Main outcome measure in Chapters 6-8 Chapter 5 outcome measure was preterm delivery therefore deliveries <24 weeks were not relevant
FETAL SEX	Missing records excluded	Potential confounder Method of identifying dichorionic twins in chapters 7-8.

4.3.3 Data Reduction - Preparing the Variables for Analyses

The next step prior to the statistical analyses of the datasets was preparing and categorising the covariates for use in the multivariate models. Most potential confounding variables were declared *a priori* as they were determined in advance to be associated with both the outcome and the exposure. Potential effect modifiers were also pre-specified in each of the analyses. The categorisation and derivation of new variables differed slightly for each of the studies in **Chapters 5-8** according to the data available and the completeness of the data. The main categorisations that were performed in each of the studies are detailed in **Table 4-2**. The cut-off points were pre-specified and chosen on the basis of previous studies and with the application of clinical knowledge. The aim was to define categories within which there was relatively little variation in the risk or event rate. In the same way, using previous studies and applying clinical knowledge, the reference range for each of the categorical variables was determined. For example, the risk in each of the 5 deprivation quintiles used in **Chapters 6-8**, with 1 being more affluent with therefore the lowest risk of perinatal death and 5 being the most deprived with the highest risk of perinatal death; there was therefore no clinical reason to collapse this variable further.

Some variables with a large number of categories were further collapsed into a smaller number of groups (for example outcome of pregnancy had up to 8 different categories and this was collapsed into three categories of livebirth, stillbirth or neonatal death). This was mainly to aid interpretation of the results. Occasionally the data were collapsed into a smaller number of groups because of data sparsity within a number of groups (data sparsity approximated as <10 events per category), for example in **Chapter 6** using the AMND, gestational age was further categorised from individual weeks to gestational week categories (<32 weeks, 33-36 weeks, 37-38 weeks and >38 weeks). This was especially true in this chapter as time-to-event survival analysis was used.

Numerical variables were converted to categorical variables to ease interpretation in the multivariable analysis and to allow for stratification by subgroup.

In **Chapter 5**, gestational age was recorded in the SMBR in days as opposed to weeks in the other registries, it was therefore treated as a continuous variable in the analyses to maintain as much information as possible.

Table 4-2: Categorisation of the main exposure, outcome and confounding variables Chapters 5-8.

VARIABLE	CATEGORISATION
MATERNAL AGE	<20, 20-29, 30-40 and >40 years
PARITY	para 0 or para ≥ 1
PERIOD OF BIRTH	1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, 2006-2010 and 2011-2015
GESTATION AND SEX SPECIFIC BIRTH WEIGHT CENTILES	<3, 3-10, 11-90, 90-97, >97 (**Chapters 7** and 8 only)
SMOKING	Smoking at booking or non-smoker
MATERNAL HEIGHT	<150, 150-154, 155-159, 160-164, 165-169, 170-174 and >175cm
GESTATIONAL AGE	34, 35, 36, 37, 38, 39, >40 weeks (in **chapter 5** maintained as a continuous variable) (in **Chapter 6** further collapsed to <32, 33-36, 37-38, >38)

4.4 Missing Data

4.4.1 Identification and Assessment of Missing Data

Along with the potential coding errors and resulting measurement bias involved in analysing routinely collected data, deciding how to with missing values is the other

key methodological challenge. Missing data can lead to bias in the study and this depends on why the data are missing. Missing values for each variable of interest in each of the datasets were all investigated and firstly assessed to determine why they were missing and classified according to the standard nomenclature; missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR)(Lunt 2011). Generally, for MCAR and MCR the chance of a data point being missing is not related to the missing data and therefore as it is random it is likely safe to remove the data point without resulting in bias. If, however the data is 'missing not at random (MNAR)' where the missing value is dependent on another variable then removal of these data points could introduce bias; for example, a variable such as body mass index (BMI) has only been recorded from 1990 onwards therefore the 'missing' records will be all from before 1990 and will not be missing at random. Removing the missing values could therefore result in bias in the results if the exposure or outcome changes over time or the relationship between them changes over time.

In **Chapters 5-6** the process of dealing with missing data was a combination of excluding the records with key variables missing as part of the data cleaning process (Table 4-1) or excluding the variable from the multivariate models if there was a lot of missing data (for example maternal smoking in **chapters 6** and **8**).

The main issue with removal of records with missing data is that this reduces the sample size and hence the study power available for the multivariate analyses making the confidence intervals wider around the estimate thus reducing the precision of the study. The option of excluding the variable with missing data from the multivariate analyses also has limitations as it results in this potential confounder not being adjusted for in the multivariate models and can lead to residual confounding of the estimate. The other method of dealing with missing values is to impute the missing data. This technique was learned towards the end of the PhD and was carried out in **Chapter 7**. The methods involved in multiple imputation are described below.

4.4.2. Multiple Imputation of Missing Values

Multiple imputation is the process of allowing for the uncertainty of missing data by creating several plausible imputed datasets and combining the results obtained for each dataset (Sterne et al. 2009). The missing values are replaced by imputed values which are sampled from their predictive distribution based on the observed data (Sterne et al. 2009). Multiple imputation was used in **Chapter 7** to impute the missing values of a number of covariates. In this Chapter multiple imputation was employed because it allowed analysis of all the potential recorded confounders in the maternity datasets whilst maintaining the largest possible sample size to estimate the effect. The ICE module in STATA was used to create the chained equations using all covariates and outcomes (Royston 2007). The number of datasets imputed was calculated based on the fraction of missing data (Allison 2012) and for **Chapter 7,** 30 imputed datasets were created.

The process of multiple imputation involved firstly assessing how much missing data was present using the 'mvpatterns' command in STATA MP, version 14.1 (stata corporation). A dry run was then performed to see how the ICE command imputed the variables and this was done using all of the other variables that were included in the multivariate analyses. The distributions of each of the imputed variables was checked against the observed values to ensure there were no impossible values imputed and ensure the distribution was similar. This was analysed by plotting histograms of imputed and observed values. Categorical variables were imputed by creating indicator variables (1 for each level within the categorical variable) so that ICE did not treat them as continuous predictors.

Once the imputed datasets were created they were analysed separately and the estimates combined by the 'mim' command before the regression analyses commands in STATA MP, version 14.1 (stata corporation)(Van Buuren, Boshuizen and Knook 1999). This command uses Rubin's rules to take account of the variability in results between the imputed datasets and the uncertainty of the missing values (Rubin 2004).

4.5 Statistical Analysis

4.5.1 Baseline Summary Statistics

Following data cleaning, preparation of variables for the analyses and assessment of missing values in **Chapters 5-8**, the characteristics of the cohorts were examined.

The first step was to examine the distributions of each of the variables to gain an understanding of the characteristics of the study population and ensure the correct outliers had been removed in the data cleaning process. Demographics were generally examined in **Chapters 5-8** separately in relation to the exposure (e.g whether individuals delivered at week 24-32 weeks, 33-36 weeks, 37-38 weeks or ≥39 weeks as per **Chapter 6**). To investigate potential confounders, the relationship between each potential confounder and the outcome/exposure was assessed using the chi squared test of association for categorical variables and chi squared test for trend for ordinal data. The chi-squared test was used to compare the observed numbers in each group with the expected frequencies if the null hypothesis (of no difference between the two groups) were true. The chi-squared test for trend was used for ordered exposures to assess the differences among the proportions in the different exposure groups and determine if there was an increasing or decreasing trend over the exposure categories (Kirkwood and Sterne 2010). Having determined that the covariate was associated with both the exposure and outcomes using summary statistics it was then decided if it was a potential confounder or not based on the associations and the clinical/biological plausibility.

4.5.2 Univariable Analyses

Following the production of baseline summary statistics, the next stage was to perform univariable analysis to determine the crude association between the exposure and the outcome (**Chapters 5-8**). The univariable analyses presented were the

unadjusted odds ratio (**Chapters 5, 7** and **8**) or the unadjusted hazard ratio (**Chapter 6**). The definition of the odds ratio used throughout this thesis is shown in the equation below:

$$Odds\ Ratio = Odds\ in\ the\ Exposed \div Odds\ in\ the\ Unexposed$$

Univariable analyses were also performed to determine the variables associated with the outcome and therefore the ones which should be considered for inclusion in the multivariable model as potential confounders (variables associated with both the exposure and outcome but not on the causal pathway). Assessment of potential confounders was performed using stratification or Mantel-Haenszel odds ratio calculations (combining the estimates from the separate strata). Assessment of how much the crude estimate had changed when the other variables were adjusted for was performed to decide as to whether confounding was present and which variables were therefore to be included in the multivariate models.

4.5.3 Multiple Linear Regression

In **Chapter 5** multiple linear regression was used to model the relationship between gestational age at delivery and a number of known risk factors for preterm birth. Gestational age in days was treated as a numerical continuous outcome variable in the multiple linear regression. Unadjusted and adjusted regression coefficients were produced and converted to ORs through the linear regression command lm() in R (version 3.4.1)(Grömping 2006). 95% Confidence intervals were calculated to determine the precision of the OR estimates. The main exposure in **Chapter 5** was maternal geographical residence during pregnancy. The effect of 296 municipalities of maternal residence were assessed and therefore a Bonferroni adjustment to the p values was calculated. The Bonferroni correction is an adjustment to p values when multiple statistical tests are being applied simultaneously to the same dataset to reduce the chances of false positive results (here accounting for 296 different municipalities of maternal residence)(Napierala 2012).

4.5.4 Multiple Logistic Regression

In **Chapters 7 and 8** multiple logistic regression modelling was used to analyse the binary outcome variables of stillbirth, neonatal death and the composite outcome of perinatal death. The logistic regression models were fitted on the log scale and exponentiated to be expressed as OR and 95% confidence intervals using the logistic() command in STATA MP version 14.1. The models took the simple form shown below:

$$Log\ (odds\ of\ disease) = a + (b\ x\ exposure)$$

Where, a = intercept (log[odds of disease in the unexposed]) and b = slope (log[odds ratio]).

ORs of variables with more than two levels were calculated relative to the referent group. The referent group was pre-determined and usually represented the most common group.

4.5.4 Time to Event Survival Analyses

In **Chapter 6** survival analysis methodology was used with gestation as the timescale and perinatal death the event. Survival analysis was chosen because it allows individuals to be followed up for different lengths of time and provides a method to test survival between two different groups. Cox regression was chosen in **Chapter 6** because it allows the rates in exposed and unexposed to vary over time with the assumption that the ratio of the rates remains constant over time (proportional hazards assumption)(Kirkwood and Sterne 2010). In Cox regression risk sets (everyone in the cohort at the time of an event) are used to generate instantaneous rates or hazards at the time of an event (number of events divided by the number of people in the cohort at that time). The ratio of the instantaneous rates is the hazard ratio (HR) and

this is the output provided from the stcox() command in STATA version 14.1 displayed in **Chapter 6**.

The Kaplan-Meier method was used to estimate cumulative survival probabilities in the twin infants according to their chorionicity status and also their assisted reproduction technologies status. Survival curves were generated in the form of a stepped line with the cumulative survival dropping with each event. The log rank test was then used to formally test the significance of the difference between the survival probabilities for each group (ie the difference been monochorionic and dichorionic twins and the difference between twins conceived by ART and those naturally conceived). The logrank test calculates the number of expected deaths in the exposed and unexposed groups and compared the expected number against the observed numbers and the p value derived from the X^2 distribution with $p < 0.05$ suggesting that the difference in survival between the two groups is statistically significantly different.

A competing risk analysis was performed in **Chapter 7** to assess the difference between the risk of stillbirth and the risk of neonatal death for deliveries by week of gestation. Competing risk analysis is a form of survival analysis used in a situation when an individual can experience more than one event and the occurrence of one hinders the occurrence of other type of events (Pintilie 2006). For example, in **Chapter 7**, death by stillbirth precludes the occurrence of death in the neonatal period hence it is a competing risk. A downside of cox proportional hazards modelling is that it ignores the competing risk and measures the 'pure' effect. In **Chapter 7**, a simple risk difference was calculated to provide a direct measure of benefit or harm in that week with a difference greater than zero favouring delivery of the baby compared to remaining *in utero*.

4.5.5 Analysis of Correlated Data

The twin data analysed in **Chapters 6-8** was recognised to be correlated, and this lack of independence between the two twin babies had to be accounted for in the analysis. Failure to account for the clustering of exposures in the twins can affect the precision around the effect estimates, because standard errors are too small leading to confidence intervals that are too narrow, and a p value that is too small.

In **Chapter 6** robust standard errors (RSE) were estimated to account for the clustering of the twins within the cox regression models. RSEs are derived using the observed variability in the data and the residuals (observed versus expected outcomes) produced by the model. The RSEs are then used to produce the 95% CIs but do not affect the parameter estimate. This approach was undertaken as it allowed the use of Cox regression modelling in STATA.

In **Chapters 7** and **8** generalised estimating equations (GEE) logistic regression analysis was undertaken to account for the clustering of exposure within twin pairs. GEE analysis allows the calculation of RSEs to produce accurate 95% CI but also considers correlations when producing effect estimates (in **Chapters 7** and **8** the effect estimate was presented as odds ratios). GEE models estimate regression parameters that have a population average interpretation. There are 3 standard options for the correlation structure which must be assessed before the GEE analysis is performed; independence (assumes no correlation in the data), exchangeable (implies that within a cluster 2 observations are equally correlated) or autocorrelation (used for repeated measurements over time)(Hanley et al. 2003). The correlation structure cannot be assessed in the standard way using the Akaike's information criterion (AIC) or 'goodness of fit' because GEE models use quasi-likelihood theory as opposed to maximum likelihood theory for independent observations. Instead the quasi-likelihood independence model criterion (QIC) was used to assess the three correlation structures listed above and the one with the lowest trace QIC was chosen and used in the final model (Cui 2007).

4.5.6 Interactions

Assessment of effect modification/interaction was performed in **Chapters 7** and **8.** Effect modification occurs when the association between an exposure and an outcome varies according to the level of a covariate (Kirkwood and Sterne 2010). If an interaction is present the odds ratios in the different strata cannot be combined and should be presented separately. Effect modification of medically indicated deliveries was assessed in **Chapter 7** by comparing models with and without an interaction term using the likelihood ratio test and subgroup analyses undertaken where these were significant.

4.6 Systematic Review Methods

In **Chapter 9,** a systematic review is presented investigating the long-term cognitive outcomes in children up to the age of 18 years born at each gestational week of term (37-42 weeks) and late preterm (34-36 weeks) compared to term (37-42 weeks)**.**

The systematic review was conducted according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Von Elm et al. 2007). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist was also used to write the review (Moher et al. 2009). Both these statements/guidelines aim to standardise and ensure the quality and rigour of systematic reviews making them easier to interpret and compare. The PRISMA checklist consists of a 27-item checklist and focuses on the assessment of the risk of bias and also the reporting of biases including publication bias.

4.6.1 Protocol and Registration

The study protocol was registered in accordance with the PRISMA checklist with PROSPERO (The International Prospective Register of Systematic Reviews)(Booth et al. 2012) and is included in **Chapter 9**. PROSPERO is an international database

of systematic reviews and it maintains a permanent published record of systematic reviews registered at inception and, thus, reducing the risk of duplication and promoting transparency (Booth et al. 2012).

Details of the study selection, the search strategy and inclusion/exclusion criteria for studies are outlined in **Chapter 9**.

4.6.2 Assessment of the Risk of Bias in Included Studies

Two independent reviewers (myself and Kirsten MacIntosh, KM) assessed the risk of bias in each study using the Risk of Bias Assessment Tool for nonrandomized studies (RoBANS)(Kim et al. 2013). Any disagreement was resolved by discussion involving a third author (Susan Shenkin).

4.6.2.1 Assessment of Selection Bias

For each included study, selection bias was assessed by determining how the participants in the study were selected for inclusion in the study (as opposed to how they were randomly assigned as in an RCT setting).

The method was assessed as:

- Low risk of bias
  - Cohort study – intervention and control groups from the same population group with the absence of outcomes confirmed at the starting point of the study
  - Case-control study – case and control groups selected from comparable population groups with the case group clearly defined
- High Risk of bias
  - Cohort study - intervention and control groups selected from different population groups and/or the presence of outcomes not confirmed at the starting point of the study

- Case-control study – case and control groups were not selected from the comparable population groups with case definition generated by self-reporting or merging of data

- Unclear risk of bias – uncertain if the patient selection resulted in a high or low risk of bias

4.6.2.2 Assessment of Bias due to Confounding

For each included study, selection biases caused by the inadequate confirmation and consideration of confounding variables were assessed.

The methods were assessed as:

- Low risk of bias

  - Major confounding variables confirmed and considered during the design phase and adequately adjusted for during the analysis phase

- High risk of bias

  - Major confounding variables were not included or not adequately considered at the design phase or adjusted for in the analysis phase

- Unclear risk of bias – uncertain if the confounding variable selection resulted in a high or low risk of bias

4.6.2.3 Assessment of Bias Caused by Measurement of Exposure

For each included study, the review assessed the performance biases caused by inadequate measurements of exposure.

The methods were assessed as:

- Low risk of bias

  - Exposure data were obtained from trustworthy sources, such as medical records or data was obtained from structured interviews.

- High risk of bias

- Data was obtained through self-report methods or if there was evidence of interviewer bias (characteristics of the investigators affect the study results) or recall bias (respondants' degree of recall affects the study results)

- Unclear risk of bias – uncertain if the exposure measurement resulted in a high or low risk of bias

4.6.2.4 Blinding of Outcome Assessments

For each included study, the review assessed the detection biases caused by inadequate blinding of outcome assessments.

The methods were assessed as:

- Low risk of bias

  - The outcome assessments were blinded or if blinding was not present its absence was judged to have no effect on outcome measurements.

- High risk of bias

  - Blinding was not performed or incomplete and it was judged to have affected the outcome measurements

- Unclear risk of bias – uncertain whether the blinding of outcome assessments resulted in a high or low risk of bias

4.6.2.5 Incomplete Outcome Data

For each included study, the review assessed the attrition biases caused by inadequate handling of incomplete outcome data.

The methods were assessed as:

- Low risk of bias

  - There was no missing data or the missing data was not thought to be relevant to the study outcomes or the quantity of missing data was similar in both the exposure and control groups.

- High risk of bias

- The missing data was thought to have affected the study outcome due to differences between the exposed and comparison groups in cohort studies or cases and controls in case-control studies, or the absence of important measurements.

- Unclear risk of bias – uncertain whether the incomplete outcome data resulted in a high or low risk of bias

4.6.2.6 Selective Outcome Reporting

For each included study, the review assessed the reporting biases caused by the selective reporting of outcomes.

The methods were assessed as:

- Low risk of bias

  - The study protocol was available, and the pre-defined primary/secondary outcomes were described as planned and all the expected outcomes were included in the study descriptors

- High risk of bias

  - Predefined primary outcomes were not fully reported, outcomes were not reported in accordance with previously defined standards, the existence of incomplete reporting regarding the primary outcome of interest or the absence of reports on expected outcomes.

- Unclear risk of bias – uncertain whether the selective outcome reporting resulted in a high or low risk of bias

4.6.3 Narrative Synthesis

Due to the large heterogeneity of the included studies in terms of the study designs and the measures of cognition used a narrative synthesis of the results was undertaken as a meta-analysis was not appropriate. Narrative synthesis refers to an approach to a systematic review that uses words and texts to summarise and explain the findings

of the synthesis as opposed to a quantitative synthesis (Popay et al. 2006). The narrative synthesis of the systematic review involved four main elements and the results and discussion are presented in **Chapter 9**:

1. Theory building in the evidence synthesis to determine how widely applicable the findings of the review were
2. Developing a preliminary synthesis which was an initial description of the included studies to organise the studies and look at patterns across them
3. Exploring relationships within and between studies and exploring the influence of heterogeneity. This step involved looking at methodological differences between the studies, differences in the baseline populations being studied as well as social heterogeneity between the studies
4. Assessing the robustness of the synthesis which refers to the methodological quality of the primary studies included in the review and concluded with an overall assessment of the strength of the evidence available.

4.7 Appendices

4.7.1 Appendix 1

4.7.1.2 Privacy Impact Assessment

Gestation of delivery of twins – influence on perinatal mortality and morbidity and childhood educational outcomes
Draft PIA October 2015

Step one: Identify the need for a PIA

Explain what the project aims to achieve, what the benefits will be to the organisation, to individuals and to other parties.

The purpose of this work is to find out how gestation (the number of weeks of pregnancy) at delivery affects the immediate and future health of the baby in twin pregnancies. Twins are a high risk pregnancy requiring specialist obstetric care and carrying a tenfold increase in death compared to singleton pregnancies. Twins also have a 50% chance of delivering early (before 37 weeks gestation) leading to problems of prematurity in the babies including death. The rate of twin pregnancies is increasing, mainly due to the increase in in-vitro fertilisation (IVF) techniques. In 2011 in Scotland/England and Wales there were over 12,000 sets of twins accounting for 2% of all births.

In singleton pregnancies, data from our own Centre and others shows that the risk of fetal death can be reduced by inducing labour at any time from 37 weeks gestation. Although early delivery reduces the risk of fetal death, it also appears to increase the risk that the child will have special educational needs at school. Unfortunately in twin pregnancies there are is limited

research into both the effect of timing of birth on outcome for twins and the effect of early delivery on reducing rates of death.

In this study I aim to look at all the sets of twins which have been delivered in Scotland since 1980. I will determine outcomes at each week of delivery from 34 weeks in order to determine when would be the optimum timing of delivery to reduce fetal death and complications. I will also look at longer term childhood educational outcomes of twin pregnancies again comparing the gestational ages they were delivered at.

Project Aims

The following specific research aims will be in investigated in this study:

1. To determine the specific perinatal and neonatal mortality and morbidity in twins following (i) spontaneous delivery and (ii) elective delivery in the absence of maternal or fetal complications. As mentioned above, this research aim has been prompted by NICE following the multiple pregnancy guideline which acknowledges a lack of evidence to support its current recommendation on elective delivery of twins.
2. To determine the perinatal risk index in twin pregnancies according to gestation at delivery. Calculating this risk index for twin pregnancies will be useful in guiding the optimum timing of delivery of twins.
3. To determine the long term educational outcomes of twin delivery by gestational age including record of additional educational support need, type of need, examination attainment level and school leaver destination. This linkage is to be done via sex discordant twins initially and then all twins if the initial analysis concludes that this is feasible.

Benefits

The results obtained in this research will benefit a wide range of clinicians and clinical researchers. In particular I anticipate that calculating the perinatal risk index (PNRI) for twin pregnancies (which has successfully been calculated for singletons and found to be lowest at 38 weeks) will be very useful in guiding the optimum timing of delivery of twins.

Additionally and specifically I will investigate long term educational outcomes of twin deliveries by gestational age. Our group has previously done this for singletons (‘*MacKay D, Smith G, Dobbie G, Pell J. Gestational age at delivery and special educational need: Retrospective Cohort Study of 407, 503 schoolchildren. PLos Medicine 2010; 7: e1000289’)* but long-term outcomes in twins have not been investigated previously. Thus the novelty of this approach will generate a major new contribution to determining the optimum timing of delivery of twins.

The use of record linkage to undertake epidemiological research into health and wellbeing from childhood to older age is an important part of the ongoing MRC strategic plan. I believe that the approaches we are taking with record linkage of maternity data to childhood outcomes will be of interest and assistance to others working in this field. In the field of obstetrics, a specialty where testing of research interventions is challenging, an issue (amongst others) which leads to lack of pharma involvement, working with large sets of patient and research data could lead to better treatment and identification of health risks and therapies, with such an approach often being the only method of addressing specific obstetric research questions.

These novel data will guide the decision making of clinicians and pregnant women and their families. Thus this research will ultimately benefit the pregnant women and their babies. In the UK over 12, 000 sets of twins are born each year, with around 1200 twin babies dying either before birth or within the first month of birth. Additionally, current advice is to offer elective delivery from 36-37 weeks of pregnancy in twins, but the risks and benefits in terms of perinatal death and the impact on long term educational outcomes are unknown. This work will also benefit twins at risk of being born prematurely, either because of elective or spontaneous preterm birth. Globally, 15 million babies are born preterm annually with over 1 million children dying each year from the complications of preterm births In the UK 55,000 babies are born preterm. Although survival rates are improving (with 77% of UK babies born at 26 weeks gestation now leaving hospital), survivors are at increased risk of short term morbidity and long term disability, including respiratory problems, motor and sensory impairment, learning difficulties, and social and behavioural difficulties. The adverse consequences of preterm birth have a major impact, not only on the babies themselves and their families, but on the health and wealth of the nation. For example in the UK alone, the complications of preterm birth result in a £2.946 billion estimated annual costs of preterm birth to the public purse in England and Wales (2006 prices). Twins have a significantly increased risk of preterm birth, and so make a major contribution to these beneficiaries.

Why is a PIA needed?

A PIA is needed because the answer to the following two questions is yes

Will information about individuals be disclosed to organisations or people who have not previously had routine access to the information?

Are you using information about individuals for a purpose it is not currently used for, or in a way it is not currently used? Yes

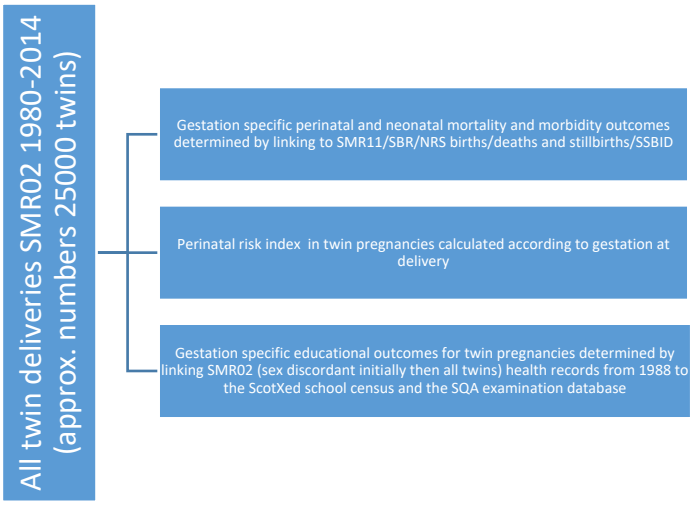
Step two: Describe the information flows

You should describe the collection, use and deletion of personal data here and it may also be useful to refer to a flow diagram or another way of explaining data flows. You should also say how many individuals are likely to be affected by the project.

I propose to carry out a population based retrospective cohort study of all twin deliveries at 34 weeks gestation or greater delivered from 1980 onwards. I will determine gestation specific perinatal and neonatal mortality and morbidity and perinatal risk index in twin pregnancies following (i) spontaneous delivery and (ii) elective delivery. I will record the following maternal and neonatal outcomes: perinatal mortality, admission to the neonatal unit, mode of delivery, postpartum haemorrhage, shoulder dystocia and uterine rupture. Confounding factors will include age, parity, deprivation category and birth weight centile for gestation. I will use multivariable logistic regression modelling to examine the relation between outcomes of elective induction of labour and expectant management for each gestation from 34 weeks. I aim to then perform this analysis again but taking out the medical indications for induction of labour (IOL) and classifying the women into 'medically indicated IOL' or 'Elective IOL' based on coding or SMR02 data.

In order to determine the longer term educational outcomes and survival following twin delivery by gestational age the SMRO2 (Maternity inpatient and Day Case – Scottish Morbidity Record) for all twin deliveries will be linked to the ScotXed school census and SQA databases at an individual level based on gender, date of birth and postcode. No names or addresses are held in the education data. The personal identifiers (names, addresses, postcodes, dates of birth and community health index – CHI) will be removed before the datasets are placed in the NHS NSS National safe haven and therefore are not accessible to me as a researcher. I will use binary and ordinal logistic regression models to explore the relationship between gestation of delivery and each outcome, adjusting for potential confounders including infant sex, maternal age and height, marital status, parity, birth-weight centile, induction of labour, mode of delivery, year of delivery, previous spontaneous and therapeutic deliveries, the 5-min apgar score and the characteristic variables from the pupil census. I will then be able to present the data in a residual and influence plot to show the trend in special educational need in twin pregnancies (both sex discordant and if the initial analysis suggests it is feasible, aggregate twins) by gestational age.

The NRS indexing team will be carrying out the indexing and linkage of personal identifiers. A three stage process will be used – firstly the matching is done (using only the linkage variables) for all twins. This will then get returned with a variable identifying 1 (sex discordant) / 0 (aggregate). The data in category 1 will be provided first for analysis and following on from this the data provided for category 0. The data linkage process is designed to protect the confidentiality of personal data to be shared. The process outlined below ensures that personal identifiers and attribute data are never transferred and stored together. Furthermore, the personal identifiers and attribute data are accessed by entirely separate teams



The NRS indexing service will be using the secure transfer protocol 'Thru' to receive personal identifiers from the Scottish Government. ISD will use the secure transfer protocol 'Globalscape' to pass the NHS numbers for the cohorts to

the NRS indexing team. Both data providers, ISD and the Scottish Government, will be using 'Globalscape' SFTP to transfer indexed content data to the National Safe Haven

Only researchers who have undertaken approved safe researcher training will be allowed to access the linked data via the National safe haven. Whilst working and analysing the data within the safe haven there is no internet access and no ability to transfer any data in a USB port/CD ROM out of the safe haven. A study folder will be created within the safe haven where results can be analysed and saved. When the analysed data is to be taken out of the safe haven it will first be checked by the research co-ordinator at eDRIS for disclosure before being transferred out of the safe haven.

The linked data will be kept in the safe haven for a period of 2 years after which will be moved to archive for the default period of 5 years. Data will be electronically deleted at the end of archiving period in the safe haven as per NSS Safe Haven policy."

Consultation requirements

Explain what practical steps you will take to ensure that you identify and address privacy risks. Who should be consulted internally and externally? How will you carry out the consultation? You should link this to the relevant stages of your project management process.

You can use consultation at any stage of the PIA process.

We will follow established processes for accessing and using health data for the purposes of the proposal. This includes the successful application to the Public Benefit and Privacy Panel for access and linkage of NHS originated data. This project will be conducted utilising the data sharing and linkage infrastructure via the electronic Data Research and Innovation Service (eDRIS). Approval has also been sought from the Scottish Government Education analytical services for access to the education data. Local R&D and ethics approval has also been approved and details of the study will be published on the HRA website.

Step three: Identify the privacy and related risks

Identify the key privacy risks and the associated compliance and corporate risks. Larger-scale PIAs might record this information on a more formal risk register.

Annex three can be used to help you identify the DPA related compliance risks.

Privacy issue	Risk to individuals	Compliance risk	Associated organisation / corporate risk
Identification of an individual through the work	Disclosure of health/education information and identity	Breach of contract and fine under DPA	Trust in Government, University and NHS services work and data collection is compromised – negative impact on future data linkage projects.
Loss of data in transfer/security breach during linkage.	Possible disclosure of identifiable information	Breach of contract and fine under DPA	As above.

4.7.1.2 Data Sharing Agreement

Introduction

The purpose of this work is to guide clinicians on the optimum timing of the elective delivery of twins. Twin pregnancy is associated with an eight to tenfold increase in perinatal mortality. Reduction in this stillbirth rate is a major priority for the NHS in both England/Wales and Scotland. Twin pregnancy is associated with significantly higher risk than singleton pregnancies. Adverse maternal outcomes associated with twin pregnancy include increased rates of pre-eclampsia, pregnancy induced hypertension, and miscarriage. Perinatal outcomes include increased rates of preterm birth, intrauterine growth restriction and stillbirth. These outcomes are of particular concern, given that the incidence of multiple births is rising. Optimising the timing of delivery is a key strategy in minimising perinatal death. It is not surprising therefore that the NICE Multiple Pregnancy Guideline Development group recommended research should be undertaken to address 'what is the incidence of perinatal and neonatal morbidity and mortality in babies born by elective birth in twin and triplet pregnancies?'.

Data from singletons illustrates the need to consider the effects of timing of delivery on short and long-term complications. In singletons appropriate timing of delivery is best determined using the perinatal risk index[3]. This is defined as the risk of death in pregnancies continuing beyond a certain gestation (compared to the risk of death in pregnancies delivered at that gestation). In singletons perinatal risk index is lowest at 38 weeks[3]. In other words, in terms of perinatal death it is safer for the baby to be delivered at 38 weeks compared to continuing the pregnancy. Consistent with this, data from our group suggests that elective induction of labour from 37 weeks in uncomplicated pregnancy reduces the risk of perinatal death[4]. These data have resulted in a Scientific Impact Paper from the Royal College of Obstetricians and Gynaecologists suggesting that early elective delivery should be offered to women with a singleton pregnancy in whom the background risk of stillbirth is high[5]. Nevertheless, in singleton pregnancies gestation at delivery has a strong, dose-dependent relationship with special educational need, with a progressive decrease in special educational needs requirement with increasing gestational age[6]. Hence although perinatal death could be reduced by early delivery in singletons, such a strategy could be associated with an increase in developmental compromise for the baby. There is almost no information on longer term outcomes in twins based on gestation at delivery.

In contrast to comprehensive data emerging for singleton pregnancy, there is little evidence on the optimum gestation of delivery of twins. NICE used data from two large population studies of fetal death in multiple pregnancy to inform their recommendation that 'in women with dichorionic twin pregnancies (two placentae and two separate chorion [monochorionic diamniotic] or shared chorion [monochorionic monoamniotic]) elective birth considered from 36 weeks'. These populations were from Japan in 1996 and London in 2001, thus neither are particularly relevant to the current UK population, although three smaller studies (n = around 5000 cumulatively) subsequently published have upheld its main conclusions. In addition to this limited evidence on gestation related perinatal mortality in twin pregnancy and the effect of early delivery, there is almost

no information on longer term childhood outcomes by gestation at delivery. This information is crucial in attempting to reduce perinatal mortality and morbidity in twin pregnancies

Organisations involved in the Data Sharing

Organisation/Business Area	*Scottish Ministers (Scottish Government)*
Information Asset Owner (IAO) (if applicable)	*Audrey MacDougall*
Operational Contact Name	*Ailie Clarkson*
Operational Contact Job Title:	*Statistician*
ICO Registration Number:	*Z4857137*

Organisation/Business Area	*University of Edinburgh*
Information Asset Owner (IAO) (if applicable)	*Professor Jane Norman*
Operational Contact Name:	*Dr Sarah Murray*
Operational Contact Job Title:	*Clinical research fellow, PhD in Reproductive Health*
ICO Registration Number:	*Z6426984*

Purpose(s) of the data sharing

Purpose(s)

The purpose of the analysis is to determine longer term educational outcomes and survival following twin delivery by gestational age. Previous studies in singleton pregnancies have demonstrated that gestation at delivery has a strong, dose-dependent relationship with special education need, with a progressive decrease in special educational needs requirement with increasing gestational age. In contract to singleton pregnancies, there is almost no information on longer term childhood outcomes by gestation at delivery in twins. The data sharing of the maternity data with the Scottish government education data is necessary to investigate the above.

Aims & Benefits

Aims

The following specific research aims will be in investigated in this study – aim 3 relates to the data linkage of ISD data to ScotXed data:

4. *To determine the specific perinatal and neonatal mortality and morbidity in twins following (i) spontaneous delivery and (ii) elective delivery in the absence of maternal or fetal complications. As mentioned above, this research aim has been prompted by NICE following the multiple pregnancy guideline which acknowledges a lack of evidence to support its current recommendation on elective delivery of twins.*
5. *To determine the perinatal risk index in twin pregnancies according to gestation at delivery. Calculating this risk index for twin pregnancies will be useful in guiding the optimum timing of delivery of twins.*
6. *To determine the long term educational outcomes of twin delivery by gestational age including record of additional educational support need, type of need, examination attainment level and school leaver destination.*

Benefits

The results obtained in this research will benefit a wide range of clinicians and clinical researchers. The research will investigate long term educational outcomes of twin deliveries by gestational age. The University group has previously carried out similar research for singletons but long-term outcomes in twins have not been investigated previously. Thus the novelty of this approach will generate a major new contribution to determining the optimum timing of delivery of twins.

These novel data will guide the decision making of clinicians and pregnant women and their families. Thus this research will ultimately benefit the pregnant women and their babies. In the UK over 12, 000 sets of twins are born each year, with around 1200 twin babies dying either before birth or within the first month of birth. Additionally, current advice is to offer elective delivery from 36-37 weeks of pregnancy in twins, but the risks and benefits in terms of perinatal death and the impact on long term educational outcomes are unknown. This work will also benefit twins at risk of being born prematurely, either because of elective or spontaneous preterm birth. Globally, 15 million babies are born preterm annually with over 1 million children dying each year from the complications of preterm births In the UK 55,000 babies are born preterm. Although survival rates are improving (with 77% of UK babies born at 26 weeks gestation now leaving hospital), survivors are at increased risk of short term morbidity and long term disability, including respiratory problems, motor and sensory impairment, learning difficulties, and social and behavioural difficulties. The adverse consequences of preterm birth have a major impact, not only

on the babies themselves and their families, but on the health and wealth of the nation. For example in the UK alone, the complications of preterm birth result in a £2.946 billion estimated annual costs of preterm birth to the public purse in England and Wales (2006 prices). Twins have a significantly increased risk of preterm birth, and so make a major contribution to these beneficiaries.

Limitations on Use

Processing of the data by University of Edinburgh must meet the conditions defined in Section 33.1 of the Data Protection Act 19989 for the processing of data for research purposes. Further, the data to be shared are to be used only for carrying out the research project described in this agreement.

Further Disclosure
There will be no onward disclosure of the individual / record level data supplied.
Some processing of the data may be carried out by NSS and the NRS Indexing Team on behalf of University of Edinburgh, who remain data controller. University of Edinburgh are responsible for ensuring that appropriate data processor agreements are in in place with subcontractors. Data will only be shared with University of Edinburgh on receipt of assurances that satisfactory agreements are in place.
Only disclosure controlled output data will be used by the project for publication and other forms of dissemination. The possibility of deductive disclosure is low given the nature and age of the data to be provided. The individuals working on the project, named in section 4.7, will provide written undertakings with the data providers not to attempt to identify individuals from the data.
The data will be accessed and analysed within the National safe haven. No individual level data or small cell data are allowed to leave the safe haven. The researcher will only be allowed to remove from the Safe Haven any data analyses performed once eDRIS has performed its robust statistical disclosure control policy to all outputs. All tables etc will be checked (and adjusted if necessary) to ensure that they do not inadvertently identify individuals.

Data to be shared

Personal Data

Personal data is defined as data which 'relates to a living individual who can be identified-

- *(a) From those data, or*
- *(b) From those data and other information which is in the possession of, or likely to come into the possession of, the data controller*

Therefore although the data available to the researcher in the safe haven is anonymised it still constitutes as personal data under condition (b). This is required in order to investigate the relationship between twin births and educational outcomes using individual pupil records linked to their corresponding health data.

Sensitive Personal Data

As per above. The data to be shared includes information as to the additional support needs of subjects. This may be considered sensitive personal data as it relates to the physical or mental health or condition of the subjects.

Data to be Matched or Linked

Data for linkage purposes

1. *Pupil census (2007-2014):*

Anonymised pupil ID
Gender
Post code
Date of birth
Payload data

Education data

1. *Pupil census (2007-2014)*

Study ID (generated by the NRS indexing team)
Census year
Stage
Free school meal (registration)
Level of English (1 – English as an Additional Language, 0 – other)
Ethnicity (standard categories)

SIMD (decile)
Student need category
Student need type
Student mainstream integration (No of half days in mainstream primary/secondary)
Student attendance at special schools/units (No of half days in special schools)
Mode of attendance at secondary school
Nature of additional support provided
Access to physical adaptation required
Access to curriculum adaptation required
Access to communication adaptation

2. *SQA Attainment Data*

Study ID (generated by NRS indexing team)
Year
Course code
Course level
Course result
SCQF level

3. *Leaver destinations*

Study ID (generated by NRS indexing team)
Year
Date of leaving (month and year only)
Destination (Initial)
Destination (follow up)

4. *Attendance, Absence and Exclusion (2007-2014, every 2 years only from 2010/11)*

Study ID (generated by NRS indexing team)
Number of half day absences, broken down by authorised, unauthorised, late, work experience.
Number of exclusions, broken down by temporary/removed from register.

5. *SMR02 (Scottish Morbidity record, Scottish birth record, National registry of Scotland births and deaths) - Data held by ISD which will be released by eDRIS for linkage to the education data*

Ethnicity – mother/child
SIMD
Carstairs dep cat
Maternal age at pregnancy
Parity
Total no of previous pregnancies
Maternal height and weight
Mother main condition and other
Presentation at delivery
Outcome of pregnancy
Duration of pregnancy
Year and month of delivery
Smoker during pregnancy
Mode of delivery
Third/forth degree tear
EBL > 1000mls
Duration of labour
Onset of labour
Prelabour membrane rupture
Induction of labour
Mother marital status
Mother diabetes
Mother drug misuse
Mother alcohol consumption
Maternal death
Stillbirth flag
Feeding intention
Feed on discharge
Baby birthweight
Baby sex
Baby discharged to
Admission to special care/neonatal unit

Year of admission/discharge to NNU
Baby main and other conditions
Problems requiring intensive care
Congenital flag

Transfer of data

The researcher will carry out a population based retrospective cohort study of all twin deliveries at 34 weeks gestation or greater delivered from 1980 onwards. Gestation specific perinatal and neonatal mortality and morbidity and perinatal risk index in twin pregnancies following (i) spontaneous delivery and (ii) elective delivery will be determined. The following maternal and neonatal outcomes will be recorded: perinatal mortality, admission to the neonatal unit, mode of delivery, postpartum haemorrhage, shoulder dystocia and uterine rupture. Confounding factors will include age, parity, deprivation category and birth weight centile for gestation. Multivariable logistic regression modelling will be used to examine the relation between outcomes of elective induction of labour and expectant management for each gestation from 34 weeks. It is then aimed to perform this analysis again but taking out the medical indications for induction of labour (IOL) and classifying the women into 'medically indicated IOL' or 'Elective IOL' based on coding or SMR02 data.

In order to determine the longer term educational outcomes and survival following twin delivery by gestational age the SMRO2 health record (Maternity inpatient and Day Case – Scottish Morbidity Record) for twin deliveries will be linked to the ScotXed school census and SQA databases at an individual level based on gender, date of birth and postcode – two files will be generated (i) sex discordant twins and (ii) aggregate twins. No names or addresses are held in the education data. The personal identifiers (names, addresses, postcodes, dates of birth and community health index – CHI) will be removed before the datasets are placed in the NHS NSS National safe haven and therefore are not accessible to me as a researcher. Binary and ordinal logistic regression models will be used to explore the relationship between gestation of delivery and each outcome, adjusting for potential confounders including infant sex, maternal age and height, marital status, parity, birth-weight centile, induction of labour, mode of delivery, year of delivery, previous spontaneous and therapeutic deliveries, the 5-min apgar score and the characteristic variables from the pupil census. This will allow the data to be presented in a residual and influence plot to show the trend in special educational need in twin pregnancies by gestational age. Given that the SG data does not have any patient names, the research will initially use sex discordance to perform the initial analysis with a plan to then use the entire twin population with both twins as an aggregate, if the initial analysis indicates that this is feasible.

The NRS indexing team will be carrying out the indexing of personal identifiers. A three stage process will be used – firstly the matching is done (using only the linkage variables) for all twins. This will then get returned with a variable identifying 1 (sex discordant) / 0 (aggregate). The data in category 1 will be provided first for analysis and following on from this the data provided for category 0, if it is agreed by both parties to this agreement that the initial analysis concludes this is feasible. The data linkage process is designed to protect the confidentiality of personal data to be shared. The process outlined below ensures that personal identifiers and attribute data are never transferred and stored together. Furthermore, the personal identifiers and attribute data are accessed by entirely separate teams

The NRS indexing service will be using the secure transfer protocol 'Thru' to receive personal identifiers from the Scottish Government. ISD will use the securely pass the NHS numbers for the cohorts to the NRS indexing team. Both data providers, ISD and the Scottish Government, will be using 'Serv-U to transfer indexed content data to the National Safe Haven

Only researchers who have undertaken approved safe researcher training will be allowed to access the linked data via a safe haven. Whilst working and analysing the data within the safe haven there is no internet access and no ability to transfer any data in a USB port/CD ROM out of the safe haven. A study folder will be created within the safe haven where results can be analysed and saved. When the analysed data is to be taken out of the safe haven it will first be checked by the research co-ordinator at eDRIS for disclosure before being transferred out of the safe haven.

The linked data will be kept in the safe haven for a period of 2 years after which will be moved to archive for the default period of 5 years. Data will be electronically deleted at the end of archiving period in the safe haven as per NSS Safe Haven policy.

Format of Data

The data will be supplied in CSV format

Frequency of transfer

Other than the initial data transfers for indexing/linkage and moving content data to the NSS safe haven as detailed in 5.1, no further data transfer is planned. Any further data transfer requests would be subject to approval of a new application to the Education Analytical Services Data Access Panel.

Access Restrictions

Only the lead researcher Dr Sarah Murray and the study supervisors – Professor Jane Norman, Professor Jill Pell and Dr Sarah Stock will have access to the data – please see section 7 for further details on how only the above access is secured.

Basis for Sharing

Legal Basis

The sharing of this data is necessary for the administration of the functions of government. The Scottish Government is a producer of Official Statistics and bound by the Statistics and Registration Services Act 2007, which establishes the Code of Practice (CoP) for Official Statistics (section 10). This requires us to:

- *Meet user needs, as defined under principle 1 of the CoP, including:*
  dissemination of official statistics to meet the requirements of informed decision making by government, public services, business, researchers and the public and to maximise public value.
- *Make statistics available to all users, as in principle 8 of the CoP, including:*
  make statistics available in as much detail as is reliable and practicable, subject to legal and confidentiality constraints and ensure that official statistics are disseminated in forms that enable and encourage analysis and re-use.

The Education (Scotland) Act 1980 and the Education (Schools) Act 1992 provide the legislation to collect data about schools and pupils (including information about the continuing education of pupils leaving a school; or employment or training taken up by such pupils on leaving).

The provision of data by EAS to others is enabled by the ScotXed School Handbook inserts, which inform pupils and parents about how the data the Scottish Government and its Local Government partners collect about pupils will be used, why it is needed and what we do to protect the information. This clarifies that individual data is used for statistical and research purposes only.

The handbook insert notes that information (including at the individual level) may be shared with partners, including Education Scotland, National Records of Scotland and academic institutions.

Further details can be found at:
http://www.gov.scot/Topics/Statistics/ScotXed/SchoolEducation/SchoolPupilCensus/SchoolHandbookInsertpupils
http://www.gov.scot/Publications/2012/09/8694
www.legislation.gov.uk/ssi/2012/130/made

Data will not be transferred outside the European Economic Area.

Data Protection Act

The data is being shared for statistical and research purposes under section 33 of the Act. The data will not be used to support measures or decisions with respect to particular individuals nor will it be used in a way that causes substantial damage or distress to any data subject.

The agreement does not extend to the sharing of data for administrative purposes through which individuals are publicly identified or have action taken directly against them as a result of data which identifies them being exchanged.

Information Assurance & Security

Personnel Security

The anonymised individual level linked data will be stored and analysed in the NSS National safe haven at the Farr Institute. The analysed data (released following disclosure control by eDRIS) will be stored on remote secure servers at the University of Edinburgh, which are password-protected and for which 'strong' passwords are in place. The information on the drive is backed-up nightly by IT support. Data for linkage will be stored on secure NSS servers.

Physical Security

The anonymised record-linked data will be stored in a safe haven. Analysed (and disclosure controlled) data will be processed on a computer within the Queen's Medical Research Institute, University of Edinburgh at Little France. This is a secure building with ID card-activated access, only available to workers based in the building. Data will be stored on remote secure servers run by the University of Edinburgh which are password protected. The drive which will be used is only accessible by the person to whom it was given and computer administrators.

The following security measures are in place at the Farr institute and also the Queen's Medical Research Institute where the required data can be accessed:

1. *Manned reception area*
2. *Swipe card access to building*
3. *No internet access allowed within the safe haven*
4. *USB ports, Cd drives and internet connections are disabled within the safe haven*

Only those individuals named in section 4.7 (who have undergone specific information governance training) will have access to the safe haven. They will not be able to extract any individual data from the National Safe Haven, output files which have been cleared for disclosure by the eDRIS coordinator will be sent to the named individuals by email.

Technical Security

The individuals named in section 4.7 have undergone training in information governance and all users of the National safe haven are required to accept the 'eDRIS user agreement' before each access to the safe haven.

eDRIS follows the Data Linkage Framework Guiding Principles including confidentiality through statistical disclosure control methods on statistical outputs, secure file transfer protocols to support transition of data between data providers and the safe haven and secure data provisioning and backup.

Management of a Security Incident

University of Edinburgh is responsible for any security incidents and should follow its established reporting processes for any incidents or loss incurred by itself or its sub-contractors (NRS Indexing Team / NSS). Should there be a security breach, Dr Sarah Murray will inform the eDRIS Research Coordinator or Information Consultant and Educational Analytical Services as soon University of Edinburgh becomes aware. Depending on the severity of the incident, this may result in the datasets being deleted.

Information Management

Freedom of Information (FOI) and Environmental Information (EIR) Requests

FoI requests for the individual information shared are generally expected to be exempt under the personal information element of the Act (section 38 and regulation 11 of EIR). University of Edinburgh is the data controller for the information shared and will be responsible for responding to FoI requests in line with the Act. University of Edinburgh will make Education Analytical Services aware when a request for the information shared is received and will seek their comments. However, the final decision on the response lies with University of Edinburgh.

Subject Access Requests (SAR)

Statistics and Research is exempt from these requests under section 33 of the data protection act.

Privacy Impact Assessment (PIA)

Completed – see attachment.

Retention & Deletion

The linked data will be kept in the safe haven for a period of 2 years after which will be moved to archive for the default period of 5 years, therefore 7 years in total. Data will be electronically deleted at the end of archiving period in the safe haven as per NSS Safe Haven policy."

The linked data will be destroyed according to the requirements of the safe haven. The researchers will have no access to patient identifiable information.

Management of Agreement

Commencement

Upon date of signature of this agreement

Duration

7 years

Review & Changes to Agreement

This agreement will be reviewed annually and any requests for changes should be made in writing, depending on the scale of the changes this may result in a new DSA being necessary.

Closure of Agreement

On closure of this agreement, all individual level datasets and any non-disclosure controlled derived data will be deleted by University of Edinburgh and their sub-contractors. The agreement will close in January 2023 or sooner at the discretion of Scottish Government, Education Analytical Services.

10. Third Party Claims

University of Edinburgh agrees to indemnify and keep indemnified the Scottish Government from and against all actions, claims, demands, liabilities, damages, losses, costs, charges and expenses (including all and any fines or monetary penalties levied by the Information Commissioner or any other regulator), interest, penalties and legal and other costs and expenses which the Scottish Government may suffer or incur in connection with or arising directly from any breach or non-performance of University of Edinburgh of any of its obligations under this Agreement or from any use or disclosure by University of Edinburgh of the data other than as allowed by this Agreement.

4.7.2 Appendix 2

4.7.2.1 Aberdeen Maternity and Neonatal Databank Approval

UNIVERSITY OF ABERDEEN

Department of Obstetrics and Gynaecology
University of Aberdeen
Aberdeen Maternity Hospital
Cornhill Road
Aberdeen, AB25 2ZL
Scotland
Tel: +44 (0)1224 438441 · Fax: +44 (0)1224 559948 · Email: sohinee.bhattacharya@abdn.ac.uk

AMND 001/16

22.01.2016

Dr Sarah Murray
University of Edinburgh
MRC Centre for Reproductive Health
Queen's Medical Research Institute,
47 Little France Crescent,
Edinburgh
EH16 4TJ

Dear Dr Sarah Murray,

Aberdeen Maternity Neonatal Databank
Access Request – "Gestation of delivery of twins – influences on perinatal mortality and morbidity and childhood educational outcomes"

I can confirm that your request for access to data from the Aberdeen Maternity Neonatal Databank has been approved by the steering committee following the assessment of the revised protocol and application form which we received on the 14th of December 2015.

The approved protocol (104490/Z/14/Z) and the version number is 4. Please notify us of any changes to the protocol in the future. The cost for data extraction is estimated as £1050 plus £1000 data access charge (Total £2050) payable once the data has been transferred.

Good luck with your research.

Yours sincerely,

Ashalatha Shetty

Dr Ashalatha Shetty
Chair – AMND Steering Committee

c.c. Katie Wilde; Alastair Coutts, Derek Turner

4.7.2.2 Regional Ethics Committee Approval

Lothian NHS Board

South East Scotland Research Ethics Committee 2
Waverley Gate
2 - 4 Waterloo Place
Edinburgh
EH1 3EG
Telephone: 0131 465 5674
Joyce.clearie@nhslothian.scot.nhs.uk
www.nhslothian.scot.nhs.uk

NHS Lothian

30 May 2018

Dr Sarah Murray
Clinical research fellow, PhD in Reproductive Health
University of Edinburgh, Centre for Reproductive Health
Queen's Medical Research Institute,
47 Little France Crescent
EH16 4TJ
Edinburgh

Dear Dr Murray

Study title:	**Gestation of delivery of twins - influence on perinatal mortality and morbidity and childhood educational outcomes**
REC reference:	**15/SS/0197**
Amendment number:	**15/SS/0197/AM01**
Amendment date:	**21 May 2018**
IRAS project ID:	**187210**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

The Sub-Committee had no significant ethical concerns regarding the amendment.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Notice of Substantial Amendment (non-CTIMP)		21 May 2018
Research protocol or project proposal	5	21 May 2018

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

NHS
Lothian

Working with NHS Care Organisations

Sponsors should ensure that they notify the R&D office for the relevant NHS care organisation of this amendment in line with the terms detailed in the categorisation email issued by the lead nation for the study.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our Research Ethics Committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

15/SS/0197:	Please quote this number on all correspondence

Yours sincerely

Mr Lindsay Murray
Chair

E-mail: joyce.clearie@nhslothian.scot.nhs.uk

Enclosures: *List of names and professions of members who took part in the review*

Copy to:

Dr Sarah Murray, University of Edinburgh

4.7.2.3 Public Benefit and Privacy Panel for Health and Social Care Approval (Chapters 7 and 8)

Public Benefit and Privacy Panel for Health and Social Care
nss.PBPP@nhs.net
www.informationgovernance.scot.nhs.uk

NHS SCOTLAND

Dr Sarah Murray
Clinical Research Fellow
University of Edinburgh
Centre for Reproductive Health
Queen's Medical Research Institute
47 Little France Crescent
Edinburgh
EH16 4TJ

Date: 07 December 2017
Your Ref:
Our Ref: 1718-0132

Dear Dr Murray

Re: Application 1718-0132-Murray/Preterm perinatal mortality in twins compared to singletons: a population study
Version: v2

Further to your conditional approval issued by the Public Benefit and Privacy Panel for Health and Social Care on 23 October 2017 I am writing to confirm that all conditions applied to the approval have now been satisfied. Your application and supporting documents have undergone proportionate governance review and have now been approved in full.

This approval is given to process data as specified in the approved application form, and is limited to this. Approval is valid for the period specified until 31 August 2019. You are required to notify the Panel Manager of any proposed change to any aspect of your proposal, including purpose or method of processing, data or data variables being processed, study cohorts, individuals accessing and processing data, timescales, technology/infrastructure, or any other relevant change.

On conclusion of your proposal, as part of NHS Scotland Governance and monitoring we will require you to complete an End of Project reporting form to demonstrate that you have complied with the obligations outlined such as data destruction or submission of references for publications of findings.

I would take this opportunity to remind you of the declaration you have made in your application form committing you to undertakings in respect of information governance, confidentiality and data protection. In particular you should be aware that once personal data (irrespective of de-identification or other controls applied) has been extracted from NHSS Board(s) and transferred to you, that you will then become the Data Controller as defined by the Data Protection Act (1998). Requests for access to NHS Scotland data as part of this approved application should be supported by evidencing a copy of your approval letter and application form to the relevant local board contacts/data providers.

Please note that summary information about your application and its approval, including the title and nature of your proposal, will be published on the panel website (www.informationgovernance.scot.nhs.uk).

I hope that your proposal progresses well,

Yours Sincerely

Ashley Gray
Panel Manager
NHS Scotland Public Benefit and Privacy Panel for Health and Social Care

Public Benefit and Privacy Panel for Health and Social Care
nss.PBPP@nhs.net
www.informationgovernance.scot.nhs.uk

NHS SCOTLAND

Dr Sarah Murray
Clinical Research Fellow
University of Edinburgh
Centre for Reproductive Health
Queen's Medical Research Institute
47 Little France Crescent
Edinburgh
EH16 4TJ

Date: 9th September 2015
Your Ref:
Our Ref: 1516-0252

Dear Dr Murray,

Re: Application 1516-0252 Murray - Gestation of delivery of twins – influence on perinatal mortality and morbidity and childhood educational outcomes

Thank you for your application for consideration by the Public Benefit and Privacy Panel for Health and Social Care. Your application has undergone proportionate governance review and has been approved.

This approval is given to process data as specified in the approved application form, and is limited to this. Approval is valid for the period specified in your application. You are required to notify the Panel Manager of any proposed change to any aspect of your proposal, including purpose or method of processing, data or data variables being processed, study cohorts, individuals accessing and processing data, timescales, technology/infrastructure, or any other relevant change.

I would take this opportunity to remind you of the declaration you have made in your application form committing you to undertakings in respect of information governance, confidentiality and data protection. In particular you should be aware that once personal data (irrespective of de-identification or other controls applied) has been extracted from NHSS Board(s) and transferred to you, that you will then become the Data Controller as defined by the Data Protection Act (1998).

Please note that summary information about your application and its approval, including the title and nature of your proposal, will be published on the panel website (www.informationgovernance.scot.nhs.uk).

I hope that your proposal progresses well,

Yours Sincerely

Nicola Starkey
(Interim) Panel Manager
NHS Scotland Public Benefit and Privacy Panel for Health and Social Care
Email: nss.PBPP@nhs.net

4.7.2.4 eDRIS user agreement

National Services Scotland (NSS)
eDRIS User Agreement

NSS is the commonly known name of the Common Services Agency (CSA)

8. REVIEW

This Agreement will be reviewed by NSS every two years or more frequently if appropriate, to take into account changes to legislation that may occur, and/or guidance from the Scottish Government, NSS and Farr Institute @ Scotland.

9. DECLARATIONS AND AGREEMENT

The parties hereto hereby declare and agree to comply with all the provisions of this Agreement as follows:-

9.1 Study Number

[illegible] 1718 – 0132

Please ensure that sections 9.2 and 9.3 are completed before returning this form to eDRIS. Where relevant section 9.4 should also be completed.

9.2 NSS eDRIS User (You)

By signing and dating below you confirm that you have read, understood and agree to comply with all the provisions of this Agreement. Any breach by you of this Agreement will result in your access being restricted and may be subject to eDRIS sanctions. NSS has a duty, and is entitled hereunder, to report legal or regulatory breaches to the appropriate authorities (such as the Information Commissioner and professional regulatory bodies).

Name: SARAH MURRAY

Position: CLINICAL RESEARCH FELLOW

Organisation: UNIVERSITY OF EDINBURGH

Signature: [signature]

Date signed: 05/05/18

Study Number: 1718 – 0132

9.3 Your Authorising Organisation
(Note: Must be signed by a Head of Department, Information Custodian, or equivalent.)

"We declare that the above named User is a bona fide researcher engaged in a reputable study for which all relevant required permissions have been granted, and that the data requested can be entrusted to this person in the knowledge that they will conscientiously discharge their obligations in regard to the confidentiality of the data. This Organisation agrees to abide by all the terms of this Agreement and shall ensure that the above named User complies with all the provisions of this Agreement.

4.7.2.5 London School of Hygiene and Tropical Medicine Approval

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636
www.lshtm.ac.uk

MSc Research Ethics Committee

Dr Sarah Murray
MSc Student
Epidemiology (DL)
LSHTM

28 June 2018

Dear Sarah ,

Study Title: Preterm perinatal mortality in boy and girl twins compared to singletons: A Scottish population study

LSHTM MSc Ethics Ref: 14968

Thank you for your application for the above MSc research project, which has now been considered by the MSc Research Ethics Committee.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application (CARE) form, and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is contingent on local ethical approval having been received, where relevant. It is the responsibility of the student and their supervisor to ensure appropriate local ethical approval is in place before a study commences (ie if you indicated this in question 40, local approval is required). Please forward confirmation of local ethics approval as soon as it is received.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Local Approval	20171117 PBPP Conditions met Letter 1718-0132	15/02/2018	V1
Local Approval	15- SS-0197 187210 Reminder_for_annual_progress_report 2017	15/02/2018	V1
Other	20171117 PBPP Conditions met Letter 1718-0132	15/02/2018	V1
Investigator CV	SM_CV	15/02/2018	V1

After ethical review

Any subsequent changes to the application must be submitted to the Committee via an Amendment form on the ethics online applications website: http://leo.lshtm.ac.uk .

Yours sincerely,

Professor Paul Milligan
Chair

MScEthics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics

Improving health worldwide

Chapter 5

Geographical Differences in Preterm Delivery Rates in Sweden: A Population-based Cohort Study

The following materials have been published in Acta Obstetrica et Gynecologica Scandinavica in 2018 (Murray et al., 2018) under the same title by Dr Sarah R Murray (SM), Mr Julius Juodakis (JJ), Mr Jonas Bacelis (JB), Dr Anna Sand (AS), Professor Jane E Norman (JN), Dr Verena Sengpiel (VS) and Professor Bo Jacobsson (BJ). JN and BJ instigated the collaboration. SM conducted the analysis of the data with input from JJ and JB and oversight from BJ, VS and JN. AS provided insight on the results of the urban versus rural analyses. SM prepared the first draft of the manuscript under the guidance of BJ and JN. All authors provided critical insight for the final draft of the manuscript. SM prepared the second draft of the manuscript following peer review and all authors approved the final version of the submitted article.

In summary, this work demonstrated that after adjustment for known risk factors of preterm birth using a number of different statistical techniques (including random and fixed effects multivariate regression and methods of adjusting for multiple testing include Bonferroni and false discovery rates) large geographical differences in the rates of preterm birth remain across countries using Sweden as a model of a very high human development index country. In a further exploratory analysis to try to identify possible environmental mechanisms to determine why these geographical differences in preterm birth rates exist gestational age appeared to be longer in several proxy measurements for urbanity (gestational age was longer in areas where there was a higher fraction of the population living in urban areas, in areas with a higher fraction of unemployment, in areas with a higher amount of built upon land and in areas with a higher number of violent crimes).

This work concluded that similar to the wide differences identified worldwide in preterm birth rates between countries there also exists wide differences within countries also exist. The finding that gestational age is longer in urban areas may offer a novel area for further research to try to unpick and understand the geographical differences in preterm birth rates within countries.

5.1 Abstract

Introduction

Preterm delivery (PTD) is a major global public health challenge. The objective of this study was to determine how the PTD rates differ throughout a country of very high human-development index and to explore rural versus urban factors.

Materials and methods

A population-based study was performed using data from the Swedish Medical Birth Register 1998 to 2013. Sweden was chosen as a model because of its validated routinely collected data and availability of individual social data. The total population comprised 1 335 802 singleton births. Multiple linear regression was used to adjust gestational age for known risk factors. A second and third model were fitted allowing separate intercepts for each municipality (as fixed or random effects). Adjusted gestational ages were converted to PTD rates and mapped to residential municipalities. Additionally, the effects of six rural versus urban factors were tested using simple weighted linear regression.

Results

The study population PTD rate was 4.1%. Marked differences from the overall PTD rate were observed (rate estimates ranged from 1.7% - 6.5%). Around 20% of the gestational age variance explained by the full model could be attributed to municipality-level effects. In addition, gestational age was found to be longer in areas with higher fraction of built upon land and other urban features.

Conclusions

After adjusting for known risk factors large geographical differences in rates of PTD remain. Additional analyses to look at the effect of environmental and socio-economic factors on gestational age revealed an increased gestational age in urban areas**.**

5.2 Introduction

Globally preterm delivery (PTD), defined as delivery before 37 weeks of gestation, remains a major public health priority and is responsible for 1.1 million neonatal deaths each year (Chang et al. 2013). As well as being the most common single cause of infant and perinatal mortality it also causes increased neonatal morbidity as it affects approximately 15 million infants worldwide (Blencowe et al. 2012). The economic burden of PTD is therefore substantial given that it affects so many babies, and is estimated that it costs the US healthcare system $26 billion yearly (Butler and Behrman 2007, Henderson et al. 2004). The WHO 'born too soon' report published in 2012 called for a 50% reduction in the mortality related to PTD in resource poor countries from 2010 - 2025 (WHO 2012). PTD rates are known to differ widely throughout the world even amongst countries with a very high human development index (ranging in 2010 from 5.3 per 100 live births in Latvia to 14.7 per 100 live births in Cyprus)(Chang et al. 2013). Sweden is a county with a very high human development index and had one of the lowest rates of PTD in 2010 (Chang et al. 2013). Why such a variation between countries exists is largely unknown and Chang *et al.* went on to conclude that the implementation of interventions (smoking cessation, cervical cerclage, progesterone, reducing unnecessary iatrogenic PTD and avoiding multiple embryo transfers) would jointly produce a relative reduction in PTD of only 5% from 9.59% to 9.07%, thus highlighting the need for substantial further research to improve etiological understanding and guide development of interventions. A recent individual participant analysis of 4.1 million births from 5 countries with a very high human development index aimed to assess the contributions of risk factors and successful interventions (Ferrero et al. 2016b). The study confirmed what has been found previously that prior PTD and pre-eclampsia were the strongest individual risk factors of PTD (Goldenberg et al. 2008, Räisänen et al. 2013) but two thirds of the cases have no attributable causes, again highlighting the urgent need for further research into the etiology of PTD. This uncertainty in etiology is reflected in the fact that the best intervention for PTD prevention is still unclear (Stock and Ismail 2016).

Environmental factors have been described as having the potential to act as 'pregnancy stressors' resulting in adverse pregnancy outcomes (Dibben and Clemens 2015). In particular, exposure to air pollution (released from dust, pollen or grinding operations) has been shown in a systematic review to increase the risk of PTD (OR 1.03, 95% CI 1.01-1.05), as has exposure to carbon monoxide exposure (Liu et al. 2003). These environmental factors differ according to geographical area, in particular with regard to rural versus urban residence with higher rates of air pollution in urban areas.

The objective of this population-based study was to use Sweden as a model of a very high human-development index country (Thérien 2012)(with one of the lowest PTD rates, an accessible public healthcare system, free antenatal care with close to 100% of the pregnant population participating and a relatively homogenous population in terms of ethnicity and socio-economic status with almost 80% of the population having intermediate or high-level education) to determine if geographical differences in PTD rates exist throughout the country. The overall singleton PTD rate in Sweden was estimated at 4.4% in 2014 by Statistics Sweden. We hypothesized that similar to the international differences (Chang et al. 2013, Ferrero et al. 2016b) wide geographical variations in PTD rates would exist. In addition, we used individual maternal and fetal risk factor adjusted gestational age to show that these differences should not be attributed to different distributions of PTD risk factors. To provide some possible causes of the observed geographic differences, we performed some further exploratory analysis on a number of environmental and socio-economic factors and gestational length.

5.3 Materials and Methods

A population-based register study was performed using data from the Swedish Medical Birth Register from 1998-2013. The mandatory Swedish Medical Birth Register collects data prospectively from the first antenatal visit and has been

maintained by the National Board for Health and Welfare since 1973. The information included in the register includes demographic data, reproductive history and complications during pregnancy, delivery and the neonatal period. All births are validated every year through individual record linkage to the Swedish Population Register, which is 99% accurate for all births in Sweden. The register is subject to quality control on an annual basis. The Swedish Medical Birth Register was complemented by linked data from Statistics Sweden to provide the individual level social data (Kernell et al. 2014). A quality analysis of the register has been previously described and it is considered to be of high quality (Cnattingius et al. 1990).

The Medical Birth Register data was merged with the maternal residence information using data from Statistics Sweden (SCB), indicating the municipality (kommun) of maternal residence at the time of pregnancy. The population of Sweden is around 10 million with 85% of the population residing in the three biggest urban areas, Stockholm, Gothenburg and Malmö. Municipality level information on median disposable household income, fraction of 16+ year old population employed, fraction of population living in urban areas, fraction of the land which is built upon and mean distance to protected nature areas were also obtained through Statistics Sweden. Information on violent crimes was obtained from the Swedish National Council for Crime Prevention.

The study period 1998-2013 was chosen as 1998 represented the introduction of the ICD-10 (International Classification of Disease) coding system. Gestational age measurement is recorded in the Swedish Medical Birth Register by best available method for each infant. This variable has been described previously and is considered to be of high quality (Morken, Källen and Jacobsson 2006) In Sweden second trimester scanning has been used since the mid 1980s onwards for gestational age measurement, which is generally regarded as the gold standard for gestational age estimation in the country. By using this study period, we could therefore be sure our measurement of the outcome of interest (gestational age) was by the best available method. Only pregnancies with an accurate gestational age measurement were

included in the study. Multiple pregnancies, stillbirths and pregnancies complicated by fetal anomalies were excluded, as these pregnancies are known to be at increased risk of PTD compared to the general population (Morken et al. 2005). The type of onset of delivery has been recorded accurately in the registry since 1991. It is currently recorded as spontaneous or induced labor, or prelabor caesarean section. Induced labor and prelabor caesarean section were classified as iatrogenic deliveries. All analyses were repeated using only spontaneous, only iatrogenic, or all deliveries together.

5.3.1 Statistical Analyses

Gestational age in days was adjusted for known individual maternal and fetal risk factors using multiple linear regression. The following variables were included in the multivariate model: maternal age at delivery (years categorized as <20, 20-29, 30-40, >40), maternal height (continuous variable), maternal smoking (categorized as non-smoker, smoking in pregnancy, smoking >10 cigarettes in pregnancy), ethnicity (binary variable categorized as Swedish born mother and other), parity (primigravida, para 1, para 2, para 3, ≥ para 4), maternal education (categorized in three levels, 1 = primary/secondary school completed, 2 = less than two years of higher education completed, 3 = at least two years of higher level education/higher degree/PhD), year of delivery, infant gender, pre-existing maternal diabetes, maternal hypertension, and gestational age measurement method. Missing covariate values were not included in the adjusted multivariate analyses. Municipality PTD rates were calculated from individual gestational age measurements in days (as percentage of deliveries at <37 weeks of gestation) and mapped across Sweden. When calculating risk-factor adjusted PTD rates, individual gestational age was replaced with the residual plus intercept from the corresponding regression model and dichotomized as above.

A funnel plot was used to demonstrate the variability in the gestational age measurements by municipality size. To show the expected distribution of estimates under the null, we calculated 95% and 99.983% (95%, Bonferroni adjusted for 296

municipalities) confidence intervals as $\mu \pm Z_{1-\alpha/2}\sigma/\sqrt{n}$, where μ and σ are country-wide estimates and Z quartile function of standard normal distribution.

A second fixed effects linear regression model was fitted, allowing separate intercepts for each municipality. The variance explained by municipality-level effects was estimated by comparing the R-squared values of the models. The analysis was then repeated using municipality as a random effect. Overall significance of the added municipality-level effects was evaluated by the F-test between the nested models.

A binomial test was performed to test the PTD rates in each municipality against the overall PTD rate in the country to determine which municipalities were statistically different from the country-wide PTD rate. The chi squared test for homogeneity was used to measure the homogeneity of PTD rates in four ways: across all municipalities, only Stockholm municipalities, only Gothenburg municipalities and only Malmö municipalities.

All analyses were undertaken using the R language (version 3.4.1). The code used for the study is available at http://github.com/Perinatal Lab/SE_MFR_GEODATA.

We went on to investigate a number of different environmental and socio-economic factors to try to understand municipality differences in PTD rates. Each factor was tested in a simple weighted linear regression setting with mean adjusted gestational age in days as the outcome. Here we used gestational age in days rather than PTD rates in order to allow us to run linear models. Weights were proportional to the number of deliveries in the municipality. The additional factors investigated were median disposable household income (thousands of Swedish crowns, data from year 2011), fraction of 16+ year old population employed (year 2010), fraction of population living in urban areas (year 2010), fraction of the land which is built upon (year 2010), mean distance from residence to protected nature areas (year 2013) and violent crimes (against life and health, rate per capita, year 2010).

5.3.2 Ethical Approval

The study was approved by the Regional Ethical Review Board in Gothenburg, Sweden (968-14). The national Board of Health and Welfare approved the use of the data from the Swedish Medical Birth Register and Statistics Sweden approved the use of the individual level social data.

5.4 Results

The total population comprised 1 554 999 singleton births in Sweden registered in the Swedish Medical Birth register between 1998-2013, of which 1 335 802 met the inclusion criteria. There were 53 713 preterm infants born in the study population, giving an overall PTD rate of 4.1%. Of the PTDs 36 356 (67.7%) were spontaneous (Figure 5-1). The maternal characteristics related to all deliveries and preterm (spontaneous and iatrogenic) deliveries are summarized in Table 5-1. Of the 53 713 mothers that delivered preterm, 88% (n = 42 247) were non-smokers (compared to 91%, n = 985 698, of all spontaneous deliveries) and 80% (n = 42 909) were Swedish born mothers (this was similar to all spontaneous deliveries, n = 863 953, 80%). 94% (n = 1 020 367) of the gestational age measurements in the cohort were by 2nd trimester ultrasound scanning, with the remainder (6%, n = 66 896) by best estimate dating by last menstrual period.

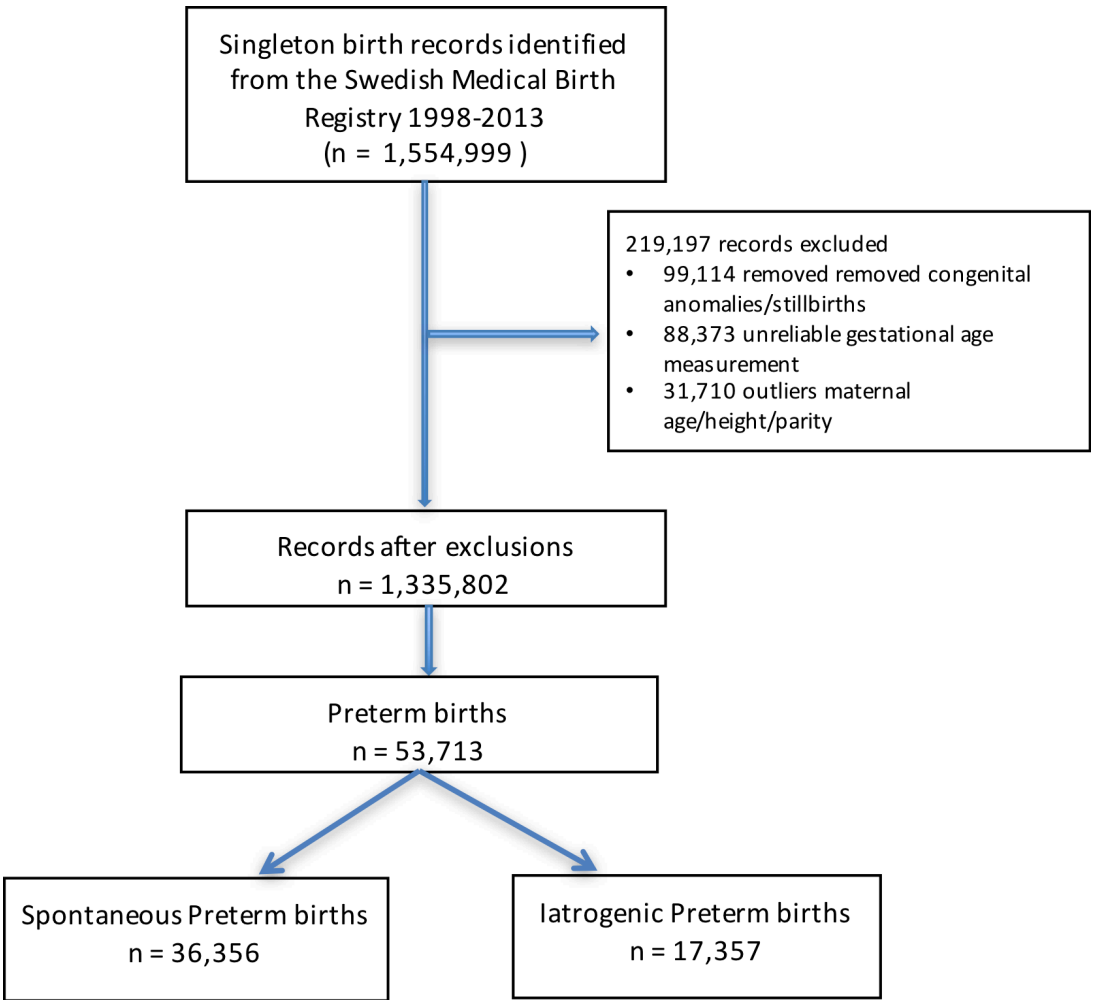


Figure 5-1: Cohort Composition

Table 5-1: Baseline demographics of all the 1,087,263 singleton deliveries and in the Swedish population 1998 – 2013.

Characteristic	N of individuals within the demographic group (% of all spontaneous births)	N PTD (% PTD from all births within the demographic group)	
		Spontaneous	Iatrogenic
Maternal age (years)			
<20	18856 (1.73)	839 (3.9)	287 (1.3)
20-30	583093 (46.27)	20343 (2.9)	8297 (1.2)
31-40	468810 (42.12)	14533 (2.4)	8209 (1.4)
>40	16504 (1.52)	641 (2.5)	564 (2.2)
Missing	0	0	0
Maternal parity			
1	480827 (44.22)	21061 (3.6)	8960 (1.5)
2	408500 (37.57)	10102 (2.0)	4609 (0.9)
3	140990 (12.97)	3383 (1.9)	2334 (1.3)
4+	56946 (5.24)	1810 (2.4)	1454 (1.9)
Missing	0	0	0
Maternal smoking			
None	985698 (90.66)	32005 (2.6)	15242 (1.3)
<10	64957 (5.97)	2736 (3.4)	1274 (1.6)
10+	22646 (2.08)	1122 (3.9)	540 (1.9)
Missing	13972 (1.29)	493 (2.8)	301 (1.7)
Maternal Diabetes			
Yes	3040 (0.28)	444 (6.3)	694 (9.8)
No	1084223 (99.72)	35912 (2.7)	16663 (1.3)
Maternal Hypertension			
Yes	3103 (0.29)	154 (2.9)	427 (7.9)
No	1084160 (99.71)	36202 (2.7)	16930 (1.3)
Maternal Ethnicity			
Swedish	863953 (79.46)	28989 (2.7)	13920 (1.3)
Non-Swedish	223310 (20.54)	7387 (2.7)	3437 (1.3)
Maternal Education			
1	248058 (22.81)	9012 (2.9)	4830 (1.6)
2	447356 (41.15)	15160 (2.8)	7176 (1.3)
3	340987 (31.36)	10489 (2.5)	4592 (1.1)
Missing	50862 (4.68)	1695 (2.8)	759 (1.2)
Method of gestational age measurement			
Ultrasound	1020394 (93.85)	33991 (2.7)	16183 (1.3)
Best estimate using last menstrual period	66869 (6.15)	2365 (2.8)	1174 (1.4)
Fetal sex			
Male	550620 (50.64)	19740 (2.9)	9021 (1.3)
Female	536643 (49.36)	16616 (2.5)	8336 (1.3)
Missing	0	0	0

PTD was strongly associated with maternal geographical residence and differed significantly throughout the country (both adjusted and unadjusted) and within major

regions (chi squared test of independence, all p values < 0.001). Figure 5-2 is a funnel plot showing the mean gestational age estimates according to municipality population size. A number of municipalities fall outside the 95% confidence area (136 out of 296 and 52 out of 296 after Bonferroni adjustment at 99.983%), indicating that they significantly deviate from the population mean. Wide variations (range 2.09% - 6.39%) in the crude spontaneous PTD rates based on municipality were observed (Appendix 1). After adjusting for potential confounding effects of maternal age, ethnicity, maternal height, smoking, parity, maternal education, baby gender, maternal hypertension and maternal diabetes, PTD rates were still widely diverse across the country (range 1.73% - 6.24%; Figure 5-3. For full area names see Appendix 1). The results of the multiple linear regression analysis are displayed in Table 5-2. Covariates accounted for approximately 1% of the variance of gestational age (R-squared 0.01); adjusted rates of spontaneous PTD were then generated which take these covariates into account. These adjusted rates were mapped across the country (and areas where they were statistically significantly different to the population mean were highlighted) and again a wide variation was observed (Figure 5-4, for full area names see Appendix 2). Unadjusted rates are shown in the appendices (Appendix 3). In supplementary analyses, we mapped spontaneous and iatrogenic rates separately (Appendix 4 and Appendix 5). The spontaneous PTD rates showed a wide variation similar to all PTD rates. When the iatrogenic rates were mapped separately across the county the differences were even larger (range 2.29 to 12.4%) and did not match areas of high spontaneous PTD rates (Appendix 5).

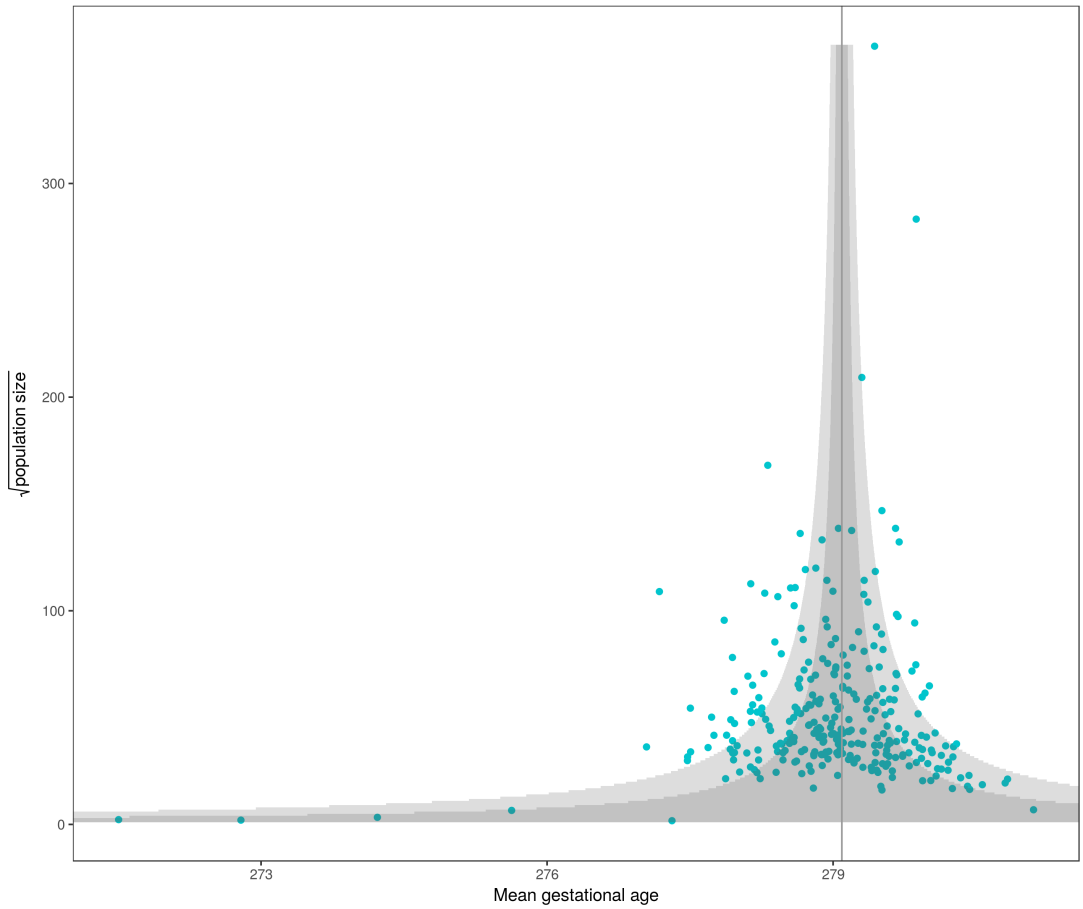


Figure 5-2: Funnel plot of gestational age and population size plotted against the population mean gestational age

Table 5-2: Results of the Multiple Linear Regression Analysis of 1 258 038 Pregnancies (77 608 removed due to missing covariates)

Variable	N	Estimate*#	SE	p value
Maternal age				
<20	17 393	-0.454	0.090	<0.0001
20-30	645 835	Ref		
31-40	571 353	0.064	0.022	<0.0001
>40	23 476	-1.011	0.078	0.004
Maternal Height	1 258 038	0.149	0.001	<0.0001
Smoking				
No smoking	1 155 513	Ref		
Smoking (<10)	75 790	-0.703	0.044	<0.0001
Smoking (>10)	26 754	-1.568	0.072	<0.0001
Parity				
Primigravida	556 768	Ref		
Para 1	469 844	-0.046	0.023	0.048
Para 2	167 782	-0.165	0.034	<0.0001
Para 3	43 503	-0.725	0.059	<0.0001
≥ para 4	20 160	-1.241	0.085	<0.0001
Ethnicity				
Swedish mother	1 044 070	Ref		
Non Swedish mother	213 987	-0.149	0.028	<0.0001
Maternal Education				
Category 1	306 215	Ref		
Category 2	539 853	0.235	0.027	<0.0001
Category 3	411 989	0.487	0.030	<0.0001
Fetal gender				
Male	642 620	-0.389	0.020	<0.0001
Female	615 437	REF		
Birth year	1 258 038	-0.032	0.002	<0.0001
Maternal diabetes				
Yes	6567	-9.379	0.142	<0.0001
No	1 251 490	Ref		
Maternal Hypertension				
Yes	5130	-4.244	0.161	<0.0001
No	1 252 927	Ref		
Gestational age measure				
USS	1 182 104	Ref		<0.0001
Other	75 53	1.011	0.043	

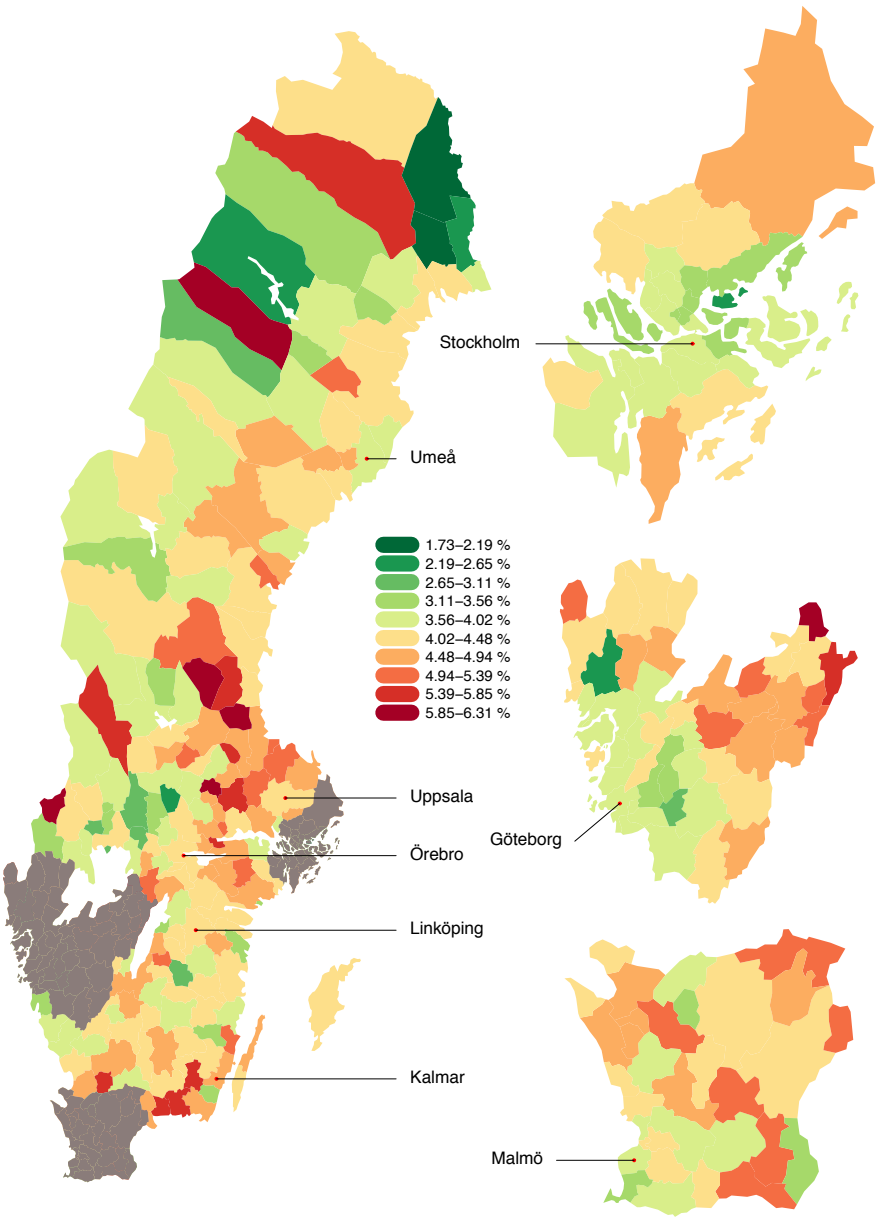


Figure 5-3: Preterm Delivery Rates Across Sweden Adjusted for Known Risk Factors from a Multiple Linear Regression Model (both spontaneous and iatrogenic deliveries are included)

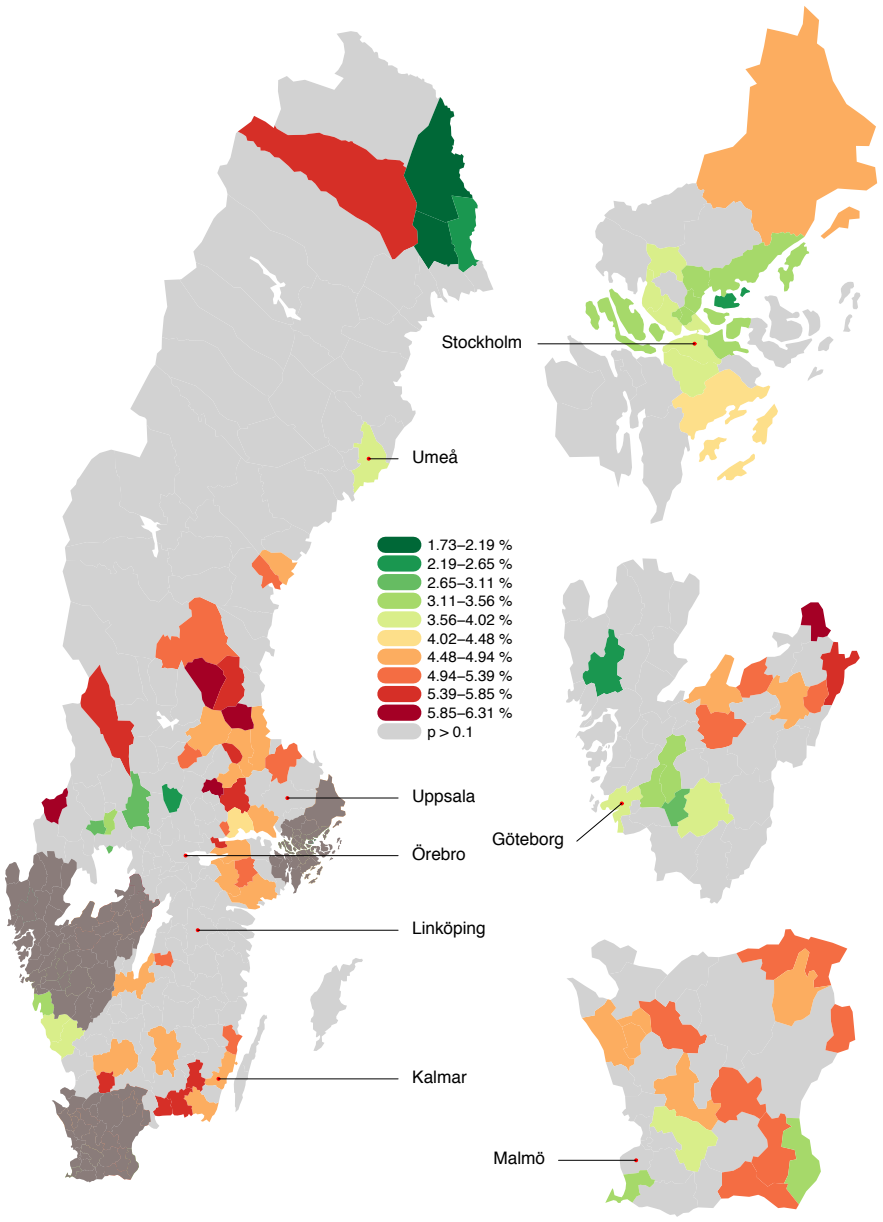


Figure 5-4: Preterm Delivery Rates Significantly Higher or Lower than the Population Mean Preterm Delivery Rate (binomial test p<0.1, no multiple testing adjustment).

When the second regression model was fitted allowing separate intercepts for each municipality the R-squared value of the model rose from 0.0127 to 0.0159 suggesting around 20% of the variance explained by the model can be attributed to municipality-level effects. The PTD rates that were significantly above and below the population

mean (using FDR adjustment for multiple testing, q-value threshold 10%) were then mapped (Appendix 6). A third model with municipality as a random effect was fitted and the estimates from the second model were almost identical therefore no further mapping was undertaken (Appendix 7). In both fixed and random effect models, addition of municipality-level effects significantly improved the model fit (F test p << 0.0001).

In the further analysis of urban versus rural environmental and socio-economic factors and their association with gestational length across Sweden, gestational age in days was significantly positively associated with several proxies for urbanity: fraction of population living in urban areas (p=0.005), fraction of population employed (p=0.02), fraction of land built upon (p<0.001) and number of violent crimes (p<0.001) (Figure 5-5). Municipality affluence (measured by median household income) was not associated with gestational age in days (p=0.99, Figure 5-5).

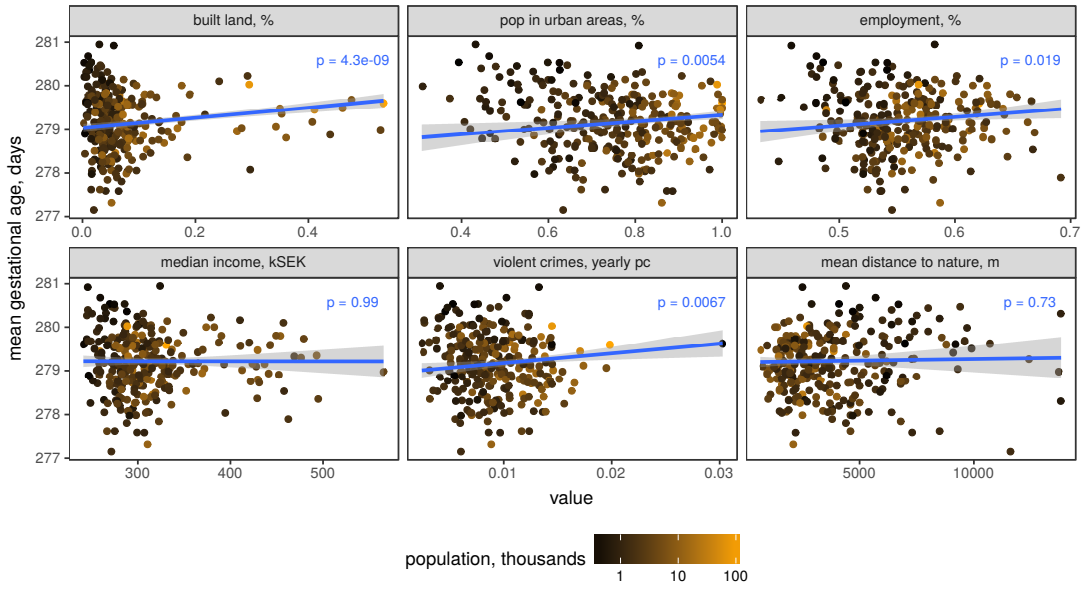


Figure 5-5: Weighted linear regression plots of environmental and socio-economic municipality features and gestational age. Points represent municipalities, weighted by their number of deliveries (n).

5.5 Discussion

This large population-based study comprising >1.3 million pregnancies reveals novel information on the association between maternal geographical residence and PTD rates. Using Sweden as a model of a very high human-development index country with one of the lowest PTD rates in the world and a public healthcare system with free antenatal care (particularly suited to this study because of its accurate and detailed routinely collected data), we have shown that PTD rates vary widely throughout the country in line with our hypothesis. Our study shows that the within country differences in PTD rates (1.40 – 5.73%) are almost as large as between country differences in PTD rates described previously (5-10% among live births in Europe (Chang et al. 2013, Blencowe et al. 2012, Zeitlin et al. 2013), and many areas of the country significantly deviate from the overall PTD rate, despite the relatively homogeneous nature of the population and health care system. This wide variation is seen for spontaneous deliveries, iatrogenic deliveries, and both categories combined. Our study highlights what is demonstrated in previous studies (Ferrero et al. 2016b, Di Renzo et al. 2011), that currently known epidemiologic risk factors for PTD account for only a small proportion of the overall variation in PTD rates, evident from the small R-squared value of the multiple linear regression model in our analysis. Our study therefore highlights further the need to consider other factors, which may be driving this association between geographical residence and PTD.

The mechanism which underlies the strong association between maternal geographical residence and PTD rates remains unclear and indeed it is surprising that a cohort of largely non-smokers (91%) and Swedish born mothers (80%) with access to one of the most comprehensive healthcare systems in the world should show such a variation in PTD rates. In our further analysis of rural versus urban environmental and socio-economic factors we have shown an association between area urbanity (and proxy measures of it such as fraction of the population employed) and a longer gestational age. The association of an increased number of violent crimes and longer gestational age is not what would be expected if it was the main stressor - more likely,

crime rate acts as a proxy measure of urbanity, which affects PTD rate through other factors. One hypothesis is that despite the public healthcare organization, more advanced healthcare practices exist within urban areas of the country. We plan to go on to look at the accessibility of specialized obstetric care providers in Sweden as previous research has showed this to be associated with pregnancy outcome(Bauer et al. 2017). The role of other environmental factors, not shown here, such as levels of sunlight (given emerging evidence about vitamin D in pregnancy and its association with a reduction in PTD rates (De-Regil et al. 2016), longitude and latitude effects and the role of water and air pollution which have previously been shown to be associated with PTD could also be investigated to further determine this urban versus rural difference (rural areas may be have increased rates of pesticide use and therefore increased water pollution and CO emissions)(Hao et al. 2015). As well as investigating environmental stressors, maternal stressors such as maternal anxiety and depression have been shown to contribute to poor obstetric outcome and increased risks of preterm birth (Ibanez et al. 2012) and are important potential confounders we were unable to address in this study.

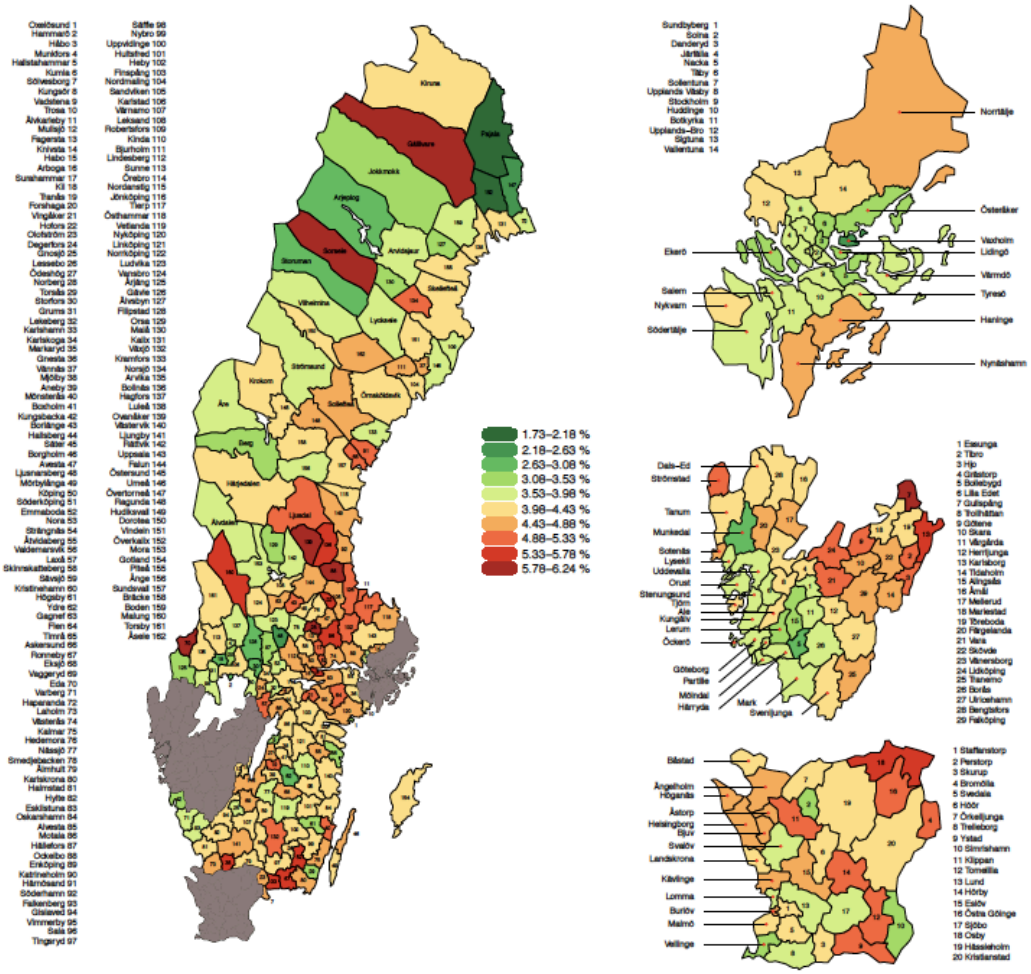
Our study has a number of strengths. Firstly, the large sample size of >1.3 million pregnancies allowed us to report the PTD rates after reliably adjusting for the known risk factors for PTD. Using a full-population database reduces the risk of selection bias and the population is homogenous with a free public healthcare system. The main strength of our study lies in the accurate measurement of gestational age and the completeness of our dataset. 94% of our gestational age measurements were by ultrasound scan and pregnancies with inaccurate gestational age measurement were excluded. Gestational age measurement is often a reason for variations in PTD rates between countries but this variation is accounted for in our analysis (Delnord, Blondel and Zeitlin 2015). Gestational age is recorded in the Swedish dataset in days and this greatly reduces the measurement noise in all our analyses. Using unique individual patient IDs we were able to accurately link the Swedish Medical birth registry data to the Statistics Sweden data with a very high match rate. Using Sweden as a model of

a very high human-development index Country we believe the results to be generalizable to other populations of similar development status.

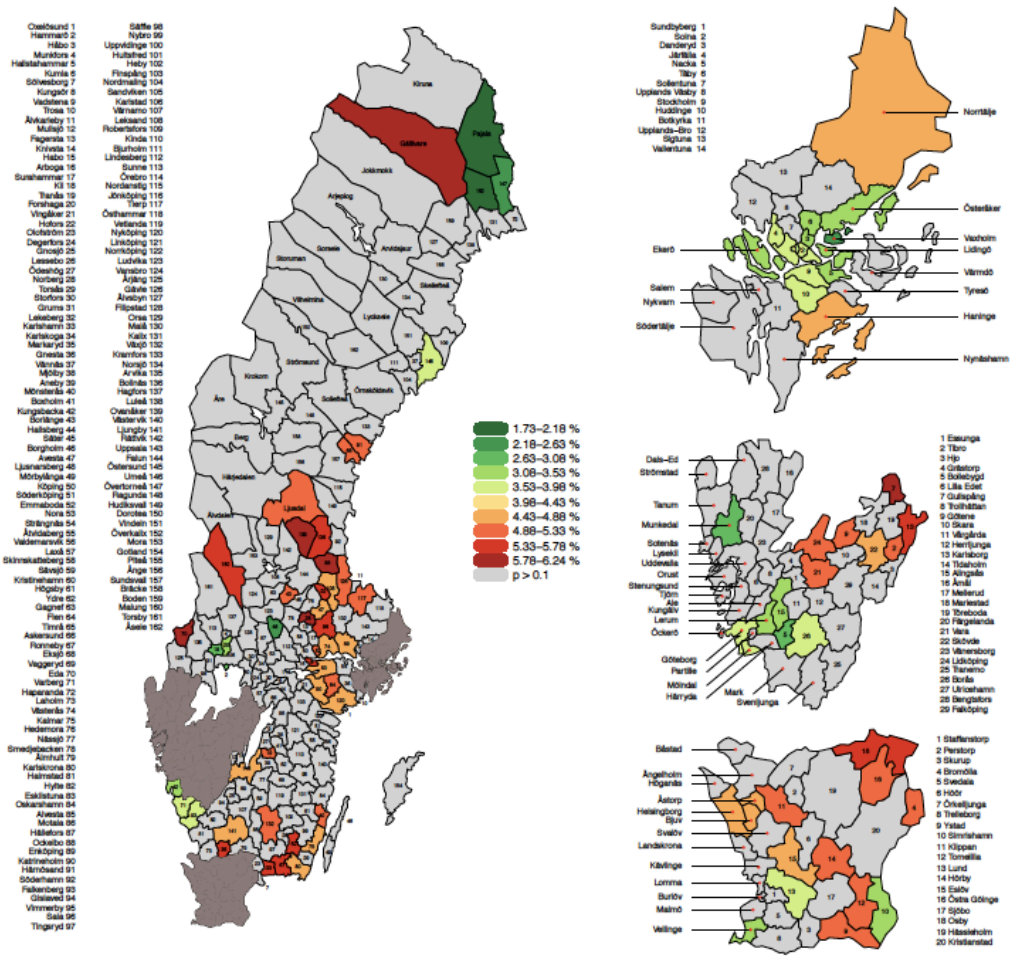
A caveat to this population-based approach is the reliance on routinely collected data method for the analysis. Large datasets are at higher risk of containing coding errors, misclassification of exposure or outcome variables, and missing data. Although we did not formally assess the Swedish Birth Register data quality for this project, it has been shown in a previous study to be 99% accurate for all births in Sweden (Morken et al. 2006). There is potential selection bias resulting from the missing covariate values in the study, which caused samples to be excluded from the adjusted analyses. However, the similarity between adjusted and unadjusted results implies that the missing data should not have a strong effect on the observed PTD rates. Another limitation is regarding the use of municipality social and environmental data from 2010 or 2013 as this data was not available for the actual year of pregnancy/birth and may have changed over the course of the study period.

In conclusion PTD rates are rising and it remains difficult to treat because of its heterogeneity and the unknown components of the etiology. Our study has shown that risk factor adjustment alone only accounts for a small amount of the variation in gestational age seen throughout a country with a very high human-development index. We observe that gestational age is longer in urban areas. We believe that future research efforts should be directed at determining the role of environmental factors and explaining the effect of urbanity on PTD rates as targeting rural municipalities may be required to reduce PTD rates.

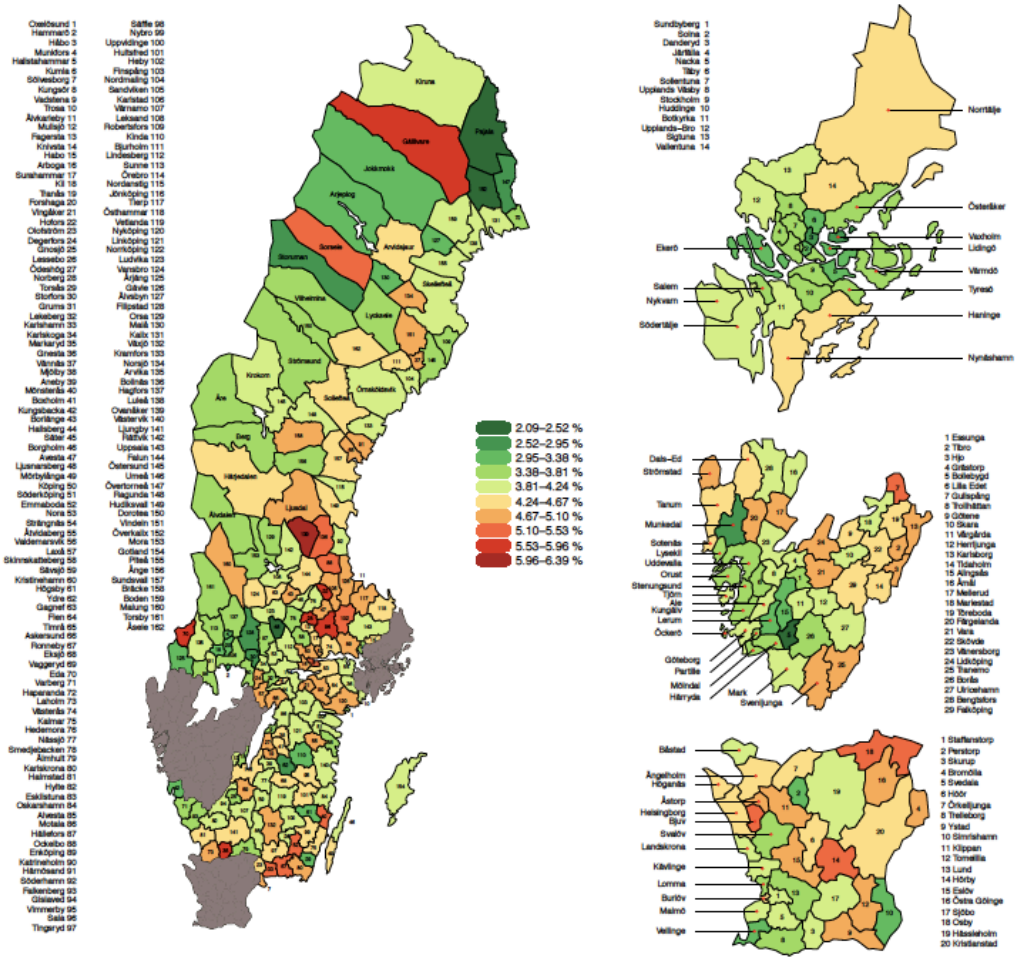
5.6 Appendices



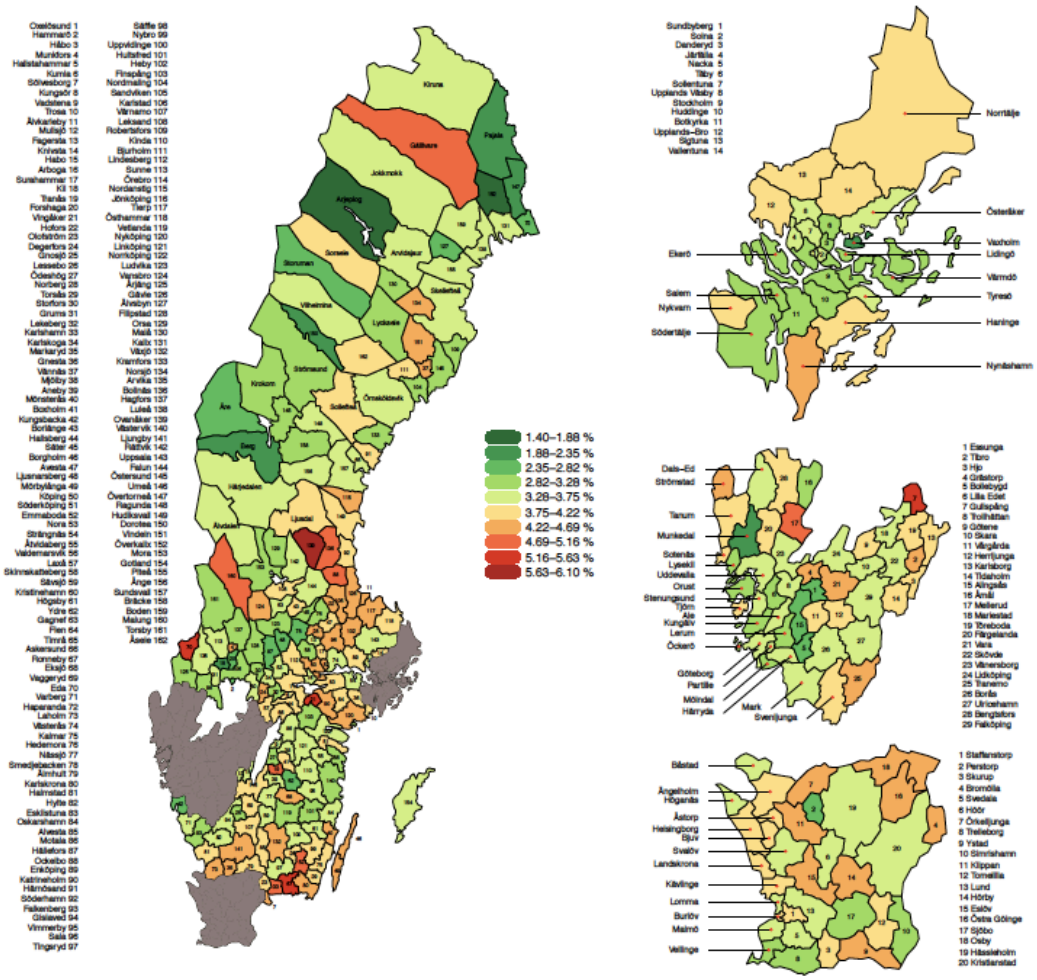
5.6.1 Appendix 1: Preterm delivery rates across Sweden adjusted for known risk factors from a multiple linear regression model (full place names).



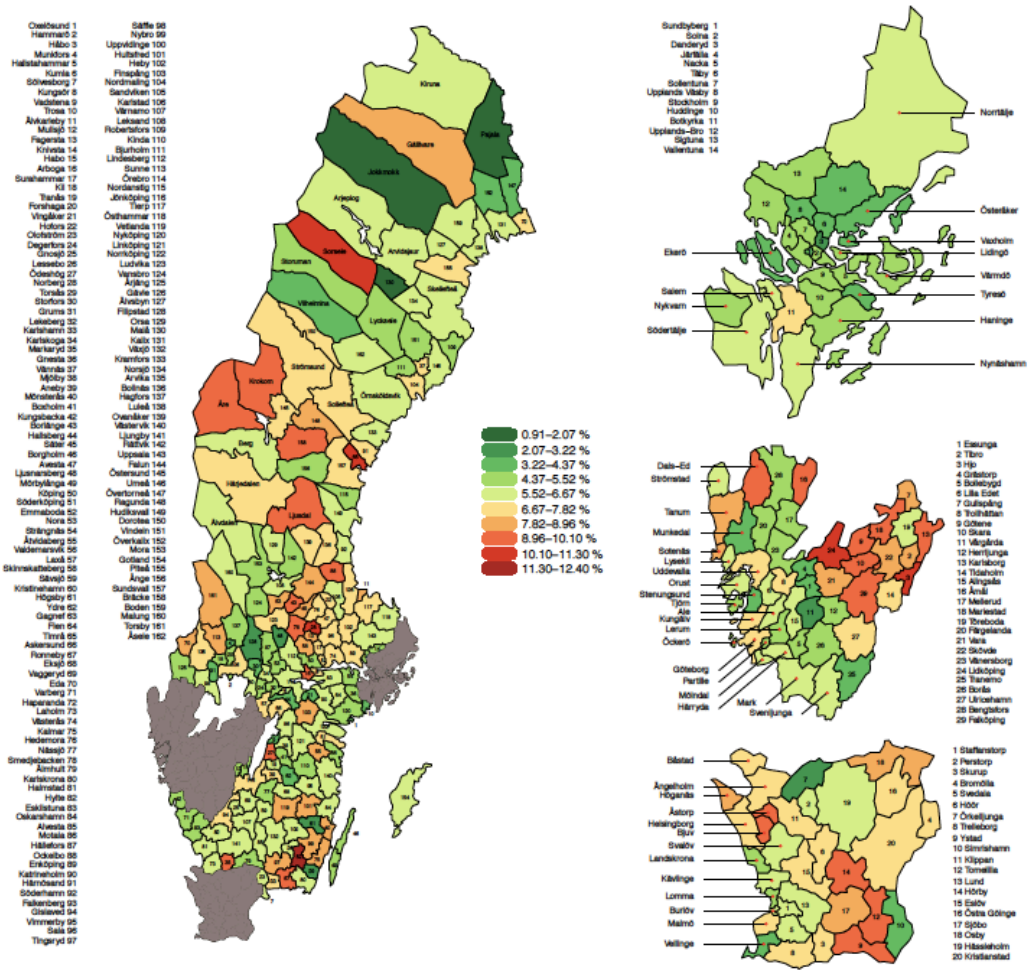
5.6.2 Appendix 2: Preterm delivery rates significantly higher or lower than the population mean preterm delivery rate (full place names).



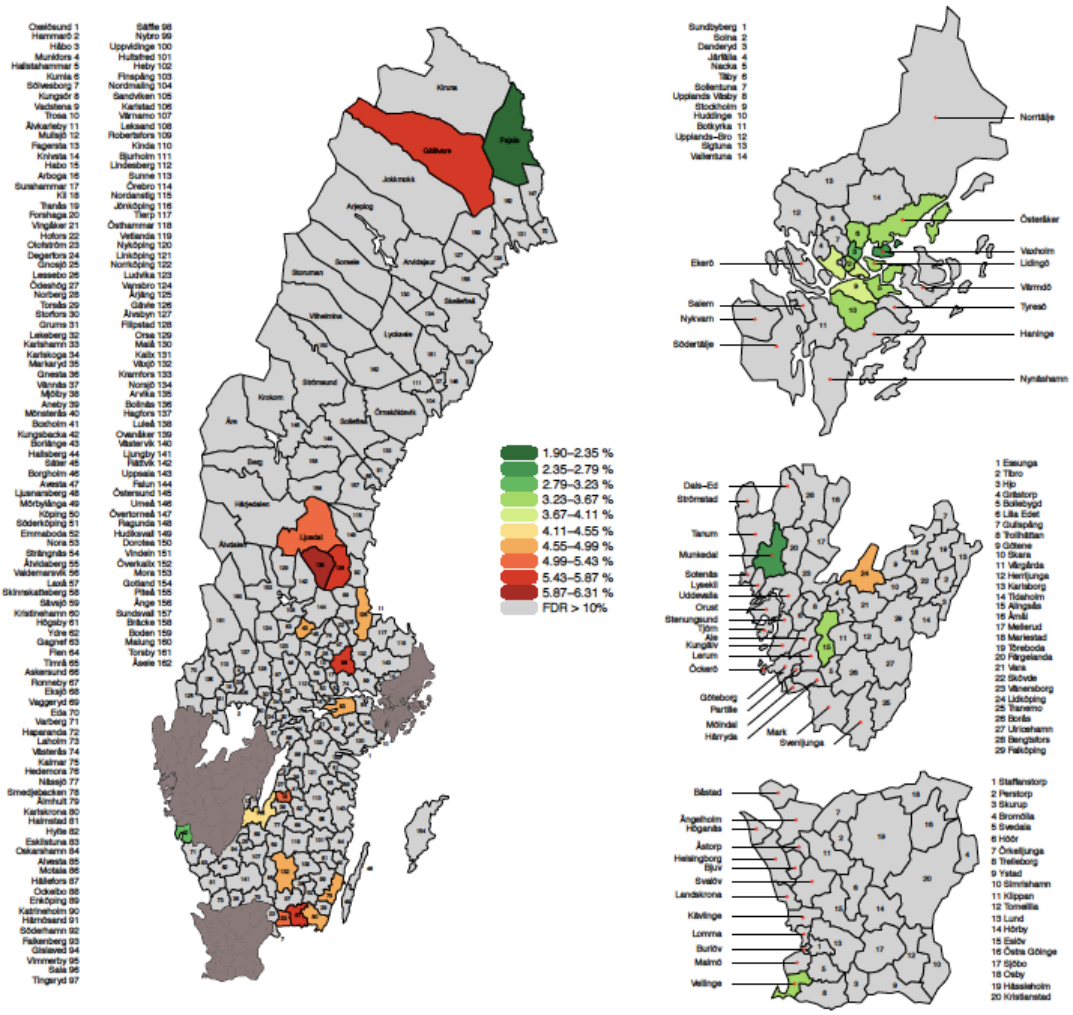
5.6.3 Appendix 3: Crude PTD rates among live births in Sweden



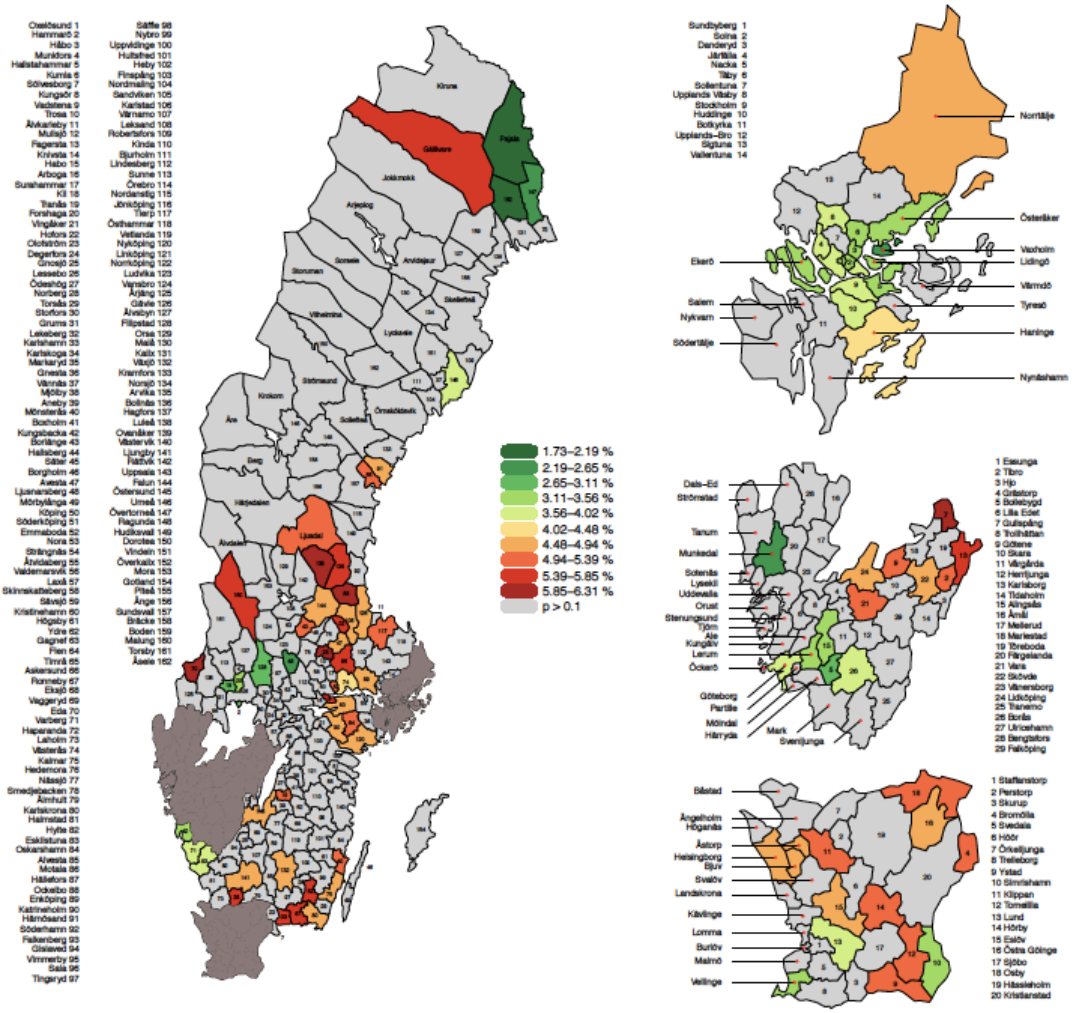
5.6.4. Appendix 4: Spontaneous PTD rates in Sweden adjusted for known risk factors.



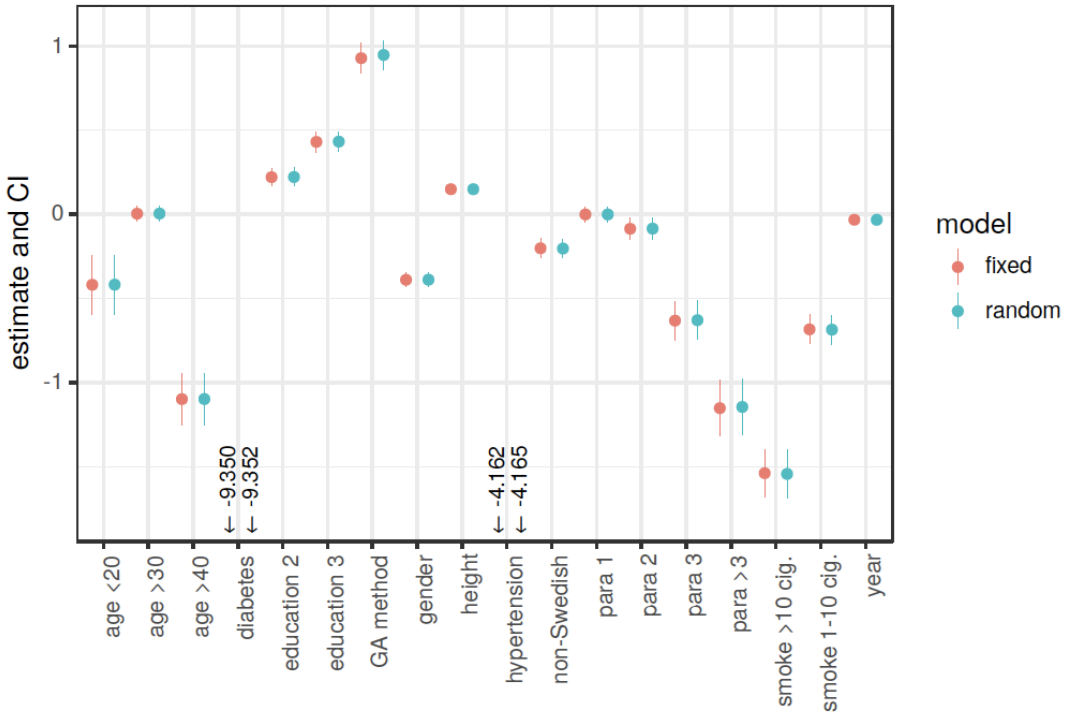
5.6.5 Appendix 5: Iatrogenic PTD rates in Sweden adjusted for known risk factors.



5.6.6. Appendix 6(a): PTD rates significantly higher or lower than the population mean PTD rate adjusted from a multivariate regression model with separate intercepts for each municipality (FDR >10%).



Appendix 6(b): PTD rates significantly higher or lower than the population mean PTD rate adjusted from a multivariable regression model with separate intercepts for each municipality ($p<0.1$).



5.6.7 Appendix 7: Effect estimates compared by method of municipality modelling (fixed and random effects).

5.7 Chapter Conclusion

The work presented in **Chapter 5** suggests that based on maternal postcode wide geographical differences in preterm birth rates exist across the country of Sweden after adjusting for a wide range of known risk factors for preterm birth. The analysis has observed that gestational age is longer in urban areas highlighting an area to focus future work on to try to determine which such large geographical differences exist within countries.

Chapter 5 explored the use of routine data to study population geographical differences in preterm birth across a county and resulted in some new observations regarding differences in urban and rural rates to guide future work. Building on the principles and methods of analysing population data gained in **Chapter 5, Chapter 6** moved onto to using population data to study outcomes in twin pregnancy.

Chapter 6

Gestational Age at Delivery of Twins and Perinatal Outcomes: A Cohort Study in Aberdeen, Scotland

The following materials have been published in Wellcome Open research in 2019 (Murray et al., 2019) under the same title by Dr Sarah R Murray (SM), Dr Sohinee Bhattacharya (SB), Dr Sarah Stock (SS), Professor Jill Pell (JP) and Professor Jane Norman (JN). JN and SB instigated the collaboration. SM conducted the analysis of the data with input from SB and oversight from JP, SS and JN. SM prepared the first draft of the manuscript under the guidance of SB and JN. All authors provided critical insight for the final draft of the manuscript and approved the final submitted article.

In summary, this work demonstrated that after adjustment for potential confounders delivery of twins at 37-38 weeks was associated with the lowest risk of perinatal death. Delivery beyond 37 weeks had a 2-fold increase in perinatal death, as did delivery at or beyond 39 weeks. In the subgroup analysis investigating chorionicity there was a 2-fold increase in perinatal death in monochorionic twins compared to dichorionic twins. In the subgroup analysis of conception status (ART conceived or naturally conceived twins) the outcomes in twins conceived through ART were the same as naturally conceived twins with no increased risk of perterm delivery or perinatal death.

This work concluded that in line with the current UK recommendations regarding the antenatal management of twin pregnancy, delivery at 37-38 weeks is associated with the lowest risk of perinatal death. It was not possible to determine a difference in perinatal death in individual weeks of gestation in dichorionic and monochorionic twins due to the small sample size in each week of gestation but similar to previous studies and in line with UK guidance recommending increased surveillance of monochorionic twin pregnancies, perinatal death was 2-fold higher in monochorionic twins compared to dichorionic twins. The finding that perinatal outcomes in twins

conceived through ART are the same as naturally conceived twins suggests that they should be managed according to the current UK guidance and this information can be used by clinicians when planning antenatal clinical management and advising families with a twin pregnancy.

6.1 Abstract

Background: Twin pregnancy is associated with a threefold increase in perinatal death compared to singletons. The objective of this study was to determine the risk of perinatal death in twins by week of gestation and to quantify the effect of known risk factors.

Methods: A cohort analysis was performed using data from the Aberdeen Maternity and Neonatal Databank (AMND). The exposure was gestational age at delivery and the primary outcome was perinatal death. Adjusted hazard ratios (aHRs) for perinatal death according to gestational age at delivery were determined by multivariate Cox proportional hazards regression with robust standard errors to account for clustering in the twin infants. Confounders and risk factors quantified and adjusted for in the model included maternal age, smoking, parity, marital status and year of birth. Kaplan-Meier time to event analysis was used to determine the differences in survival according to chorionicity and assisted reproduction technologies (ART) conception status.

Results: The population comprised of 7,420 twin babies born between 1950 and 2013 in the Grampian area of Northern Scotland. There were 272 stillbirths in the cohort (3.67%) and 273 neonatal deaths (3.68%). Compared to delivery at 37-38 weeks, delivery at or beyond 39 weeks was associated with a significant increase in perinatal death (aHR 2.00 [95% CI 1.45-2.78]). Monochorionic twins had a 2-fold increase in perinatal death compared to dichorionic twins (aHR 2.15, 95% CI 1.60-2.90). Twins conceived by ART did not have a greater risk of perinatal death compared to those naturally conceived (aHR 1.21, 95% CI 0.87-1.68)

Conclusion: This study suggests that delivery of twins at 37-38 weeks is associated with the lowest risk of perinatal death.

6.2 Introduction

Twin pregnancies have a threefold greater perinatal death rate overall compared to singleton pregnancies (Manktelow et al. 2014). The larger perinatal mortality is thought to be due to the greater preterm birth rates in twins with approximately 50% of twins delivering preterm (<37 weeks gestation) compared to 6% of singletons (ISD Scotland 2009). The gestation with the lowest absolute perinatal death rate is earlier in twins compared to singletons (Doss et al. 2012). However, delivery before term in singletons has been shown to be associated with an increased risk of neonatal morbidity. Hence risks of perinatal and neonatal mortality and morbidity have to be balanced when making decisions regarding timing of delivery of twins (Peter et al. 2013).

Despite accounting for only 3% of all live births, multiple pregnancies have a threefold higher economic burden on healthcare systems compared to singleton pregnancies because of the increased caesarean sections and neonatal unit admissions (RCOG 2017). Due to increases in assisted reproduction technologies (ART) in recent years the twin birth rate has risen and is set to continue to rise.

Optimising the timing of delivery is a key strategy in reducing perinatal death and morbidity in twin pregnancies. UK clinical guidelines support a policy of elective delivery from 37 weeks in dichorionic pregnancies (two placentae and two separate chorions) and 36 weeks in monochorionic pregnancies (one placenta and either one or two chorions) (Visintin et al. 2011) in order to reduce adverse short term outcomes in twins such as perinatal mortality. This strategy is informed by data from epidemiological studies on gestational age specific outcomes: however, these studies lack detail about the accuracy of the pregnancy dating, fail to adjust for the clustered outcomes of twin pregnancies and lack information on chorionicity, which is a key risk factor for adverse pregnancy outcome. Monochorionic twins have a perinatal mortality rate of 11.6% compared to 5% in dichorionic twins (Kilby 2017).

Chorionicity is therefore a very important factor to consider when attempting to determine optimum timing of delivery of twins.

Randomised controlled trials investigating optimum timing of delivery in twins have not been adequately powered to assess perinatal death (Dodd et al. 2012, Suzuki et al. 2000). A recent systematic review using data prospectively collected from randomised controlled trials, and therefore different from observational studies, had findings in line with the current UK practice recommending elective delivery from 37 weeks in dichorionic and 36 weeks in monochorionic twins to minimise perinatal deaths (Cheong-See et al. 2016). This review lacked information on whether the twins were conceived by ART procedures or were naturally conceived. This is important because ART pregnancies (twins and singletons) often have additional obstetric risk factors such as advanced maternal age and nulliparity (McDonald et al. 2005, Joy, McClure and Cooke 2008). Despite good evidence that singleton pregnancies conceived by ART procedures are at increased risk of obstetric and perinatal complications (Pandey et al. 2012), the evidence on pregnancy outcomes of twins conceived by ART procedures is conflicting. A systematic review and meta-analysis demonstrated no differences in perinatal outcomes between twins conceived by ART and those naturally conceived (Helmerhorst et al. 2004) however some studies showed increased rates of caesarean section and small for gestational age in the twins conceived by ART procedures compared with naturally conceived twin pregnancies (Bernasko et al. 1997).

This aim of this study was to explore the relationship between gestation at delivery and perinatal death in twins and determine whether this varies by chorionicity and ART conception status.

6.3 Methods

6.3.1 Study Design and Participants

We carried out a registry-based cohort study using all twin births in the Grampian area of Scotland between 1950 and 2013. Data were obtained from the Aberdeen Maternity and Neonatal Databank (AMND). The AMND has collected information on pregnancy related events in women living in Grampian since 1950; a relatively stable population with approximately 5,000 births per year. The Aberdeen Maternity Hospital (AMH) is the only maternity facility in Aberdeen city and >99% of residents deliver there (Davies, Bell and Bhattacharya 2016). The database is subject to regular quality assurance checks and completeness of the database is checked annually against the National Health Service (NHS) records. The methods used for data coding (using ICD-9) of the AMND and full details of the database have been described previously (Bhattacharya and Campbell 2005, Bhattacharya, Townend and Bhattacharya 2010, Humphrey and Tucker 2009, Ayorinde et al. 2016). The study was approved by the AMND steering committee. Individual patient consent and further ethical approval was not required as the study used secondary analyses of anonymised data.

6.3.2 Inclusion and Exclusion Criteria

Women were included if they had a twin delivery at 24 weeks' gestation or greater within the study period. Pregnancies complicated by congenital anomaly were excluded. Pregnancies were excluded if the gestational age at delivery was missing or recorded as greater than 43 weeks gestation, maternal age less than 10 years and if parity was missing or recorded as greater than 14.

6.3.3 Outcomes, Exposures and Covariates

The exposure of interest was gestational age at delivery. In the AMND this is recorded as the number of completed weeks of gestation on the basis of the estimated date of delivery recorded in the clinical record. Prior to 1985 this was calculated from the date of the last menstrual period with ultrasound scan dating thereafter. Gestational age was treated as an ordinal variable grouped into two-week periods

below 34 weeks gestation and one week from 34 weeks. The primary outcome was extended perinatal death of one or both twins defined as either antepartum/intrapartum stillbirth (infant born showing no signs of life) or neonatal death (death of a liveborn infant in the first four weeks of life). For the multivariate analyses and the analysis stratified by chorionicity (binary variable; monochorionic and dichorionic) and ART conception (binary variable; assisted conception/no assisted conception) we further categorised gestational age into the following categories due to the sparsity of events in some of the categories; <32 weeks, 33-36 weeks, 37-38 weeks [reference] and >38 weeks.

The following variables were considered to be potential confounders in the multivariate regression analyses: maternal age at delivery (categorised as <20, 21-24, 25-29, 30-34, 35-39, and >40 years), parity during the index pregnancy (binary variable categorised as para 0 or para ≥ 1), year of birth (categorised as 1950-75, 1976-2000, 2000-2013), area socioeconomic deprivation quintile of postcode of residence (defined by the Scottish Index of Multiple Deprivation [SIMD] 2012; 1 [most affluent] to 5 [most deprived] (Morris and Carstairs 1991)), maternal height (categorised as 141-150, 151-160, 161-170, and >170 cm), smoking (categorised as current smoker, ex-smoker or never-smoker), marital status (binary variable categorised as married/co-habiting or single), medically indicated induction of labour and maternal complications in pregnancy (binary variable categorised as no maternal complication or any of pre-eclampsia, hypertensive disease, diabetes or antepartum haemorrhage).

6.3.4 Statistical Analyses

Summary statistics were derived and compared by gestational age using chi squared test for categorical data and chi-squared test for trend for ordinal data. To determine the association between gestational week of delivery and the risk of perinatal death Cox proportional hazard regression modelling was used. To obtain the adjusted hazard ratios (aHR) for the effect of gestational age at delivery on perinatal death a

Cox regression model was fitted with the following covariates – maternal age at delivery, maternal parity, marital status, maternal height and maternal complications. We calculated robust standard errors to account for the clustering of twins within mothers.

Entries which had missing values were examined in the summary statistics but excluded from the univariate and multivariate analyses. Maternal smoking was not included in the final model due to the amount of missing values but a sensitivity analysis of complete cases with missing cases was performed.

In the subgroups of pregnancies with chorionicity and ART data available the analysis was repeated stratifying by each variable and the relationship to perinatal death assessed using Kaplan-Meier analyses and Cox proportional hazards models in which gestational age was the time scale and perinatal death the event (Nicolaides et al. 2016). aHRs were estimated, and time-to-event curves compared using the log-rank test.

P values for hypothesis tests were two sided and statistical significance set at $P<0.05$. All analyses were undertaken using STATA MP, version 14.1 (stata corporation).

6.4 Results

The AMND contained 7,894 records of twin infants born in Grampian over the study period of 1950-2013 of which 7,420 were eligible for inclusion in the analysis. There were 544 perinatal deaths (232 stillbirths and 312 neonatal deaths) in this cohort. Figure 6-1 displays the process of deriving the study cohort.

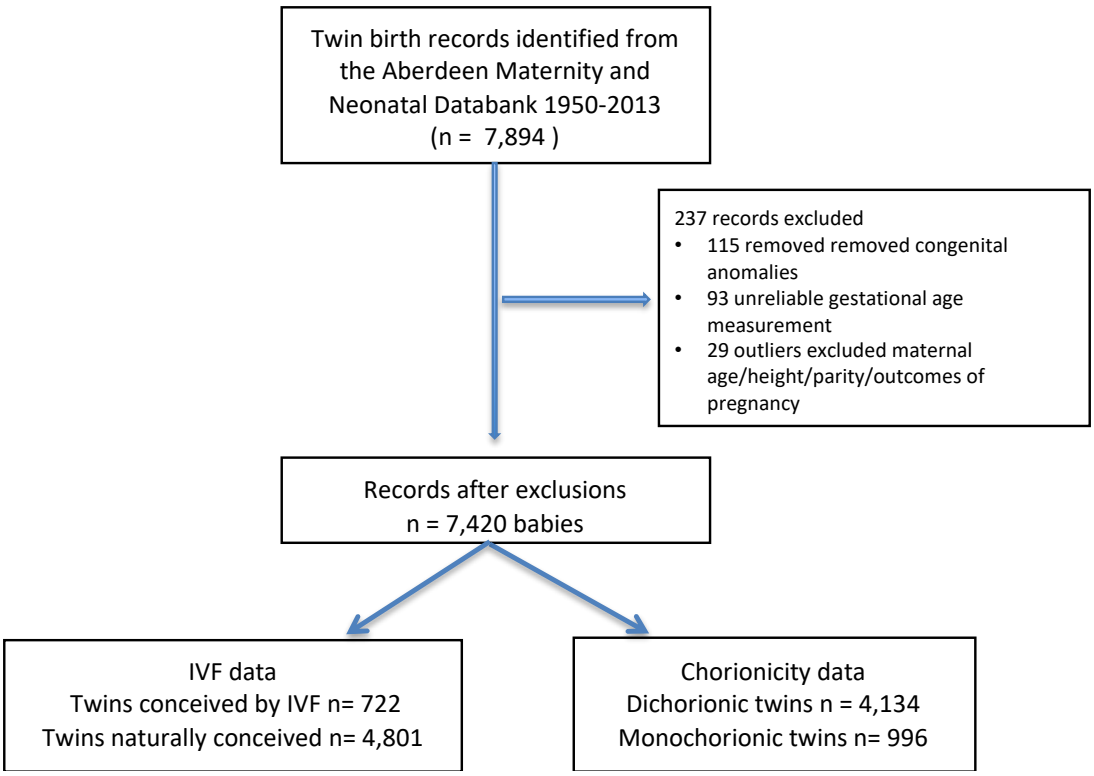


Figure 6-1: Cohort Composition

Table 6-1 summarises the pregnancy characteristics of the cohort. Among the twin infants, the largest proportion were delivered between 37 and 38 weeks (n = 2,363, 31.83%) and overall 3,615 (48.72%) delivered prematurely (<37 weeks gestation).

Table 6-1: Baseline summary statistics of the population of 7,420 twins born in Grampian, Scotland

Pregnancy Characteristic	Total N	N (%) in each gestation age group in weeks				P value
		24-32	33-36	37-38	≥39	
Perinatal death (combined stillbirth and NND[a])						
No	6876	464 (6.8)	2774 (40.3)	2299 (33.4)	1339 (19.5)	<0.001
Yes	588	276 (50.7)	101 (8.6)	64 (11.8)	103 (18.9)	
Maternal age						
15-20	486	56 (11.5)	192 (39.5)	112 (23.1)	126 (25.9)	<0.001
21-24	1290	170 (13.2)	513 (39.8)	345 (26.7)	262 (20.3)	
25-29	2272	216 (9.5)	806 (35.5)	786 (34.6)	464 (20.4)	
30-34	2108	196 (9.3)	812 (38.5)	704 (33.4)	396 (18.8)	
35-39	1066	80 (7.5)	458 (43.0)	366 (34.3)	162 (15.2)	
>40	198	22 (11.1)	94 (47.5)	50 (25.3)	32 (16.2)	
Missing	0					
Maternal smoking						
Ex-smoker	274	34 (12.4)	128 (46.7)	94 (34.3)	18 (6.6)	<0.001
Smoker	1675	192 (11.5)	640 (38.2)	473 (28.2)	370 (22.1)	
Non-smoker	3476	310 (8.9)	1486 (42.8)	1180 (34.0)	500 (14.4)	
Missing	1995	204 (10.2)	621 (31.1)	616 (30.9)	54 (27.8)	
Maternal height						
141-150	224	26 (11.6)	98 (43.8)	64 (28.6)	36 (16.1)	<0.001
151-160	2706	300 (11.1)	981 (36.3)	809 (29.9)	616 (22.8)	
161-170	3546	332 (9.4)	1408 (39.7)	1156 (32.6)	650 (18.3)	
>170	702	36 (5.1)	274 (39.0)	286 (40.7)	106 (15.1)	
Missing	242	46 (19)	114 (47)	48 (20)	34 (14)	
Year of birth						
1950-75	1957	168 (8.6)	577 (29.5)	524 (26.8)	688 (35.2)	<0.001
1976-2000	3347	368 (11.0)	1142 (34.1)	1229 (36.7)	608 (18.2)	
2001-2013	2116	204 (9.6)	1156 (54.6)	610 (28.8)	146 (6.9)	
Missing	0					
Parity						
Primigravida	3077	418 (13.6)	1335 (43.3)	856 (27.8)	468 (15.2)	<0.001
Parous	4343	322 (7.4)	1540 (35.5)	1507 (34.7)	974 (22.4)	
Missing	0					
Marital status						
Married	6034	592 (9.8)	2231 (37.0)	1961 (32.5)	1250 (20.7)	<0.001
Other[b]	1382	148 (10.7)	642 (46.5)	402 (29.1)	190 (13.8)	
Missing	4	0	2 (50)	0	2 (50)	

Maternal preconditions						
No	4337	506 (1.6)	1599 (36.5)	1366 (31.2)	906 (20.7)	<0.001
Yes	3042	234 (7.7)	1276 (41.9)	997 (32.8)	536 (17.6)	
Missing	0					
Chorionicity						
Monochorionic	996	112 (11.2)	506 (50.8)	270 (27.1)	108 (10.8)	<0.001
Dichorionic	4134	344 (8.3)	1658 (40.1)	1500 (36.3)	632 (15.3)	
Missing	2290	284 (12.4)	711 (31.0)	593 (25.9)	702 (30.7)	
Social deprivation category						
1	656	66 (10.1)	268 (40.9)	194 (29.6)	128 (19.5)	
2	1032	100 (9.7)	386 (37.4)	364 (35.3)	182 (17.6)	0.02
3	2130	214 (10.1)	797 (37.4)	717 (33.7)	402 (18.9)	
4	1330	112 (8.4)	538 (40.5)	418 (31.4)	262 (19.7)	
5	478	64 (13.4)	184 (38.5)	140 (29.3)	90 (18.8)	
Missing	1794	184 (10.3)	702 (39.1)	530 (29.5)	378 (21.1)	

6.4.1 Perinatal Outcomes According to Gestation at Delivery

Most perinatal deaths occurred in the extreme preterm period of 24-25 weeks (n = 99, 81.15%). Table 6-2 summarises the results of the univariate and multivariate Cox proportional hazards regression analyses using outcomes at 37-38 weeks as the referent. After adjusting for potential confounders, compared to delivery at 37-38 weeks, delivery at or above 39 weeks was associated with an increased risk of perinatal death (aHR 2.00, 95% CI 1.45-2.78). Delivery before 37 weeks was also associated with an increased risk of perinatal death (<32 weeks aHR 17.86, 95% CI 13.47-23.69, 33-36 weeks aHR 1.40, 95% CI 1.02-1.97). When the results were analysed by individual weeks, with 37 weeks as the referent, the relationship between perinatal death and gestation at delivery was reverse J-shaped (figure 2) with a decreasing risk of perinatal death with increasing gestational age up to 35 weeks. There was a very strong association between extreme preterm birth and perinatal death (24-25 wk: aHR 50.23 [95% CI 32.62-77.34], figure 6-2). The results were similar when we ran the Cox regression analyses for the n= 5,269 twin infants with information available on maternal smoking (24-32 weeks compared to 37-38 weeks aHR 17.48 [95% CI 12.13-25.18], 33-36 weeks compared to 37-38 weeks aHR 1.16 [95% CI 0.71-1.75] and >38 weeks compared to 37-38 weeks aHR, 3.10 [95% CI 2.07-4.77

Table 6-2: Univariate and Multivariate Cox regression analysis with robust standard errors of the association between gestational age at delivery and perinatal death in twin pregnancies (n=7,176).

Pregnancy Characteristic	N	Perinatal deaths N	Crude Hazard ratio (95% CI)	Adjusted hazard ratio (95% CI)[#]
Gestational age (wks)				
24-32	740	276	18.20 (13.87-23.89)	17.86 (13.47-23.69)
33-36	2875	101	1.40 (1.03-1.92)	1.41 (1.01-1.97)
37-38	2363	64	1	1
≥39	1442	103	2.50 (1.83-3.41)	2.01 (1.45-2.78)
Maternal Age (yrs)				
15-20	486	45	1.34 (0.95-1.89)	1.04 (0.72-1.52)
21-24	1290	126	1.42 (1.11-1.81)*	1.11 (0.87-1.41)
25-29	2272	161	1	1
30-34	2108	127	0.84 (0.66-1.07)	0.94 (0.74-1.20)
35-39	1066	72	0.95 (0.71-1.27)	1.19 (0.90-1.58)
>40	198	13	0.92 (0.51-1.65)	0.94 (0.54-1.63)
Maternal Height (cm)				
141-150	224	13	0.64 (0.36-1.13)	0.59 (0.35-1.01)
151-160	2706	239	1	1
161-170	3546	247	0.77 (0.64-0.93)*	0.88 (0.74-1.06)
>170	702	28	0.43 (0.39-0.64)*	0.74 (0.49-1.34)
Year of delivery				
1950-75	1957	220	2.16 (1.71-2.73)*	1.86 (1.38-2.52)*
1976-2000	3347	207	1.13 (0.89-1.42)	1.02 (0.78-1.33)
2001-2013	2116	117	1	1
Marital status				
Married	6034	451	1	1
Other	1382	93	0.89 (0.71-1.23)	1.02 (0.81-1.30)
Parity				
Primigravida	3077	254	1.26 (1.06-1.50)*	1.09 (0.90-1.33)
Parous	4343	290	1	1
Maternal complications				
No	4377	377	1	1
Yes	3042	167	1.62 (1.34-1.96)	1.27 (1.05-1.54)*
Medical indication for induction				
No	5730	474	1	1
Yes	1690	70	2.08 (1.62-2.68)	1.08 (0.81-1.44)

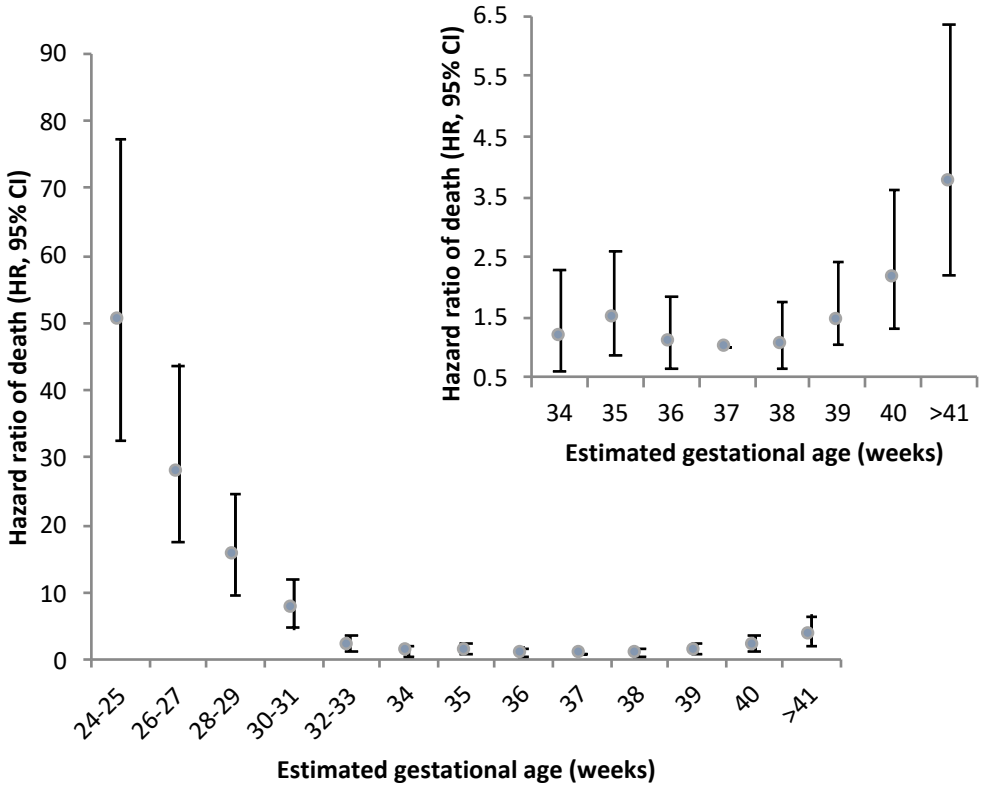


Figure 6-2: Adjusted HR of perinatal death in twins by gestation at delivery (inset data from 34 weeks onwards)

6.4.2 Perinatal Outcomes According to Gestation at Delivery Stratified by Chorionicity

Data on chorionicity was available for 5,130 twin babies, of which 4,134 (81%) were dichorionic and 996 (19%) were monochorionic (Figure 6-1). There was a highly statistically significant difference in survival between monochorionic and dichorionic twins (overall HR for death in monochorionic twins compared to dichorionic twins 2.15, 95% CI 1.60-2.90), log rank test value 37.41, $p<0.001$, Figure 6-3).

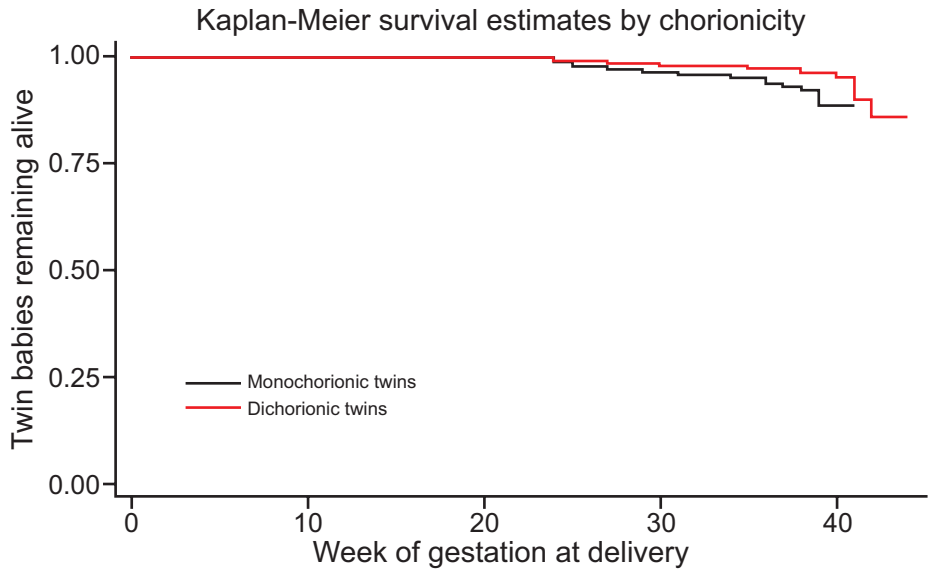


Figure 6-3: Kaplan-Meier plot of gestational age and perinatal death stratified by chorionicity

In dichorionic twin pregnancies, compared to delivery at 37-38 weeks, only deliveries <32 weeks had higher rates of perinatal death (aHR 30.14, 95% CI 17.94-50.64). Similarly, in monochorionic twin pregnancies delivery at <32 weeks was the group with a significantly higher risk of perinatal death than those delivered at 37-38 weeks (aHR 25.56, 95% CI 10.09-64.75).

6.4.3 Perinatal Outcomes According to Gestation at delivery stratified by Conception by Assisted Reproduction Technologies

Data on ART conception was available on 5,523 twin infants, of which 722 (13.07%) were conceived by ART procedures. There was no evidence of a difference in survival between ART conceived and naturally conceived twins (overall HR for perinatal death in ART conceived twins compared to spontaneously conceived twins 1.09, 95% CI 0.79-1.50, log rank test value 3.64, p=0.07, Figure 6-4).

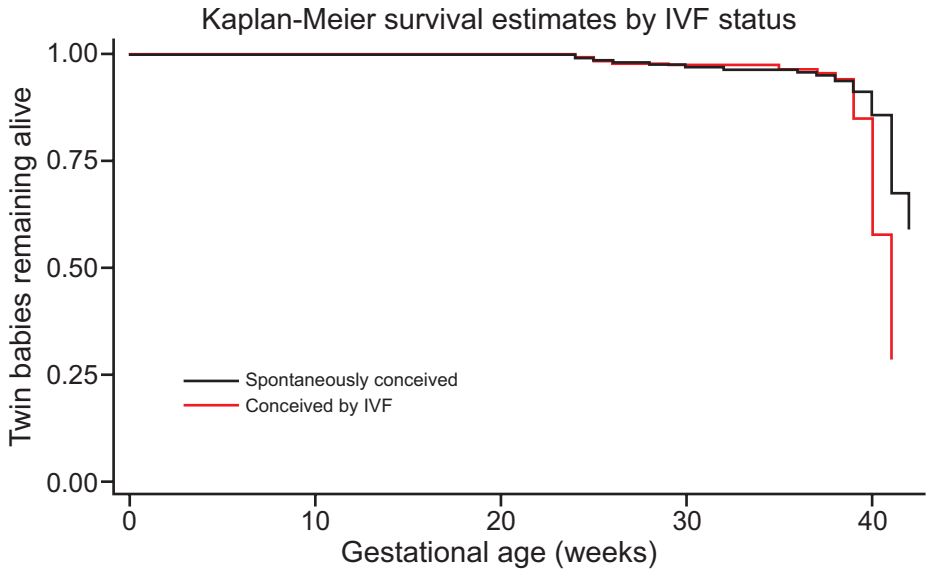


Figure 6-4 Kaplan-Meier plot of gestational age and perinatal death stratified by in vitro fertilization or spontaneous conception

ART conceived twins were also no more likely to deliver preterm compared to spontaneously conceived twins (aHR 1.02, 95% CI 0.85-1.24). In both twins conceived by ART and those twin pregnancies spontaneously conceived compared to delivery at 37-38 weeks there was an increased risk of perinatal death in deliveries <32 weeks (aHR 18.58 [95% CI 12.70-27.19] in spontaneously conceived twins and aHR 19.91 [95% CI 6.54-60.71] in ART conceived twins) and in deliveries at or beyond 39 weeks (aHR 2.91, 95% CI 1.87-4.54 in spontaneously conceived twins and aHR 19.49, 95% CI 6.60-57.59 in ART conceived twins).

6.5 Discussion

6.5.1 Main Findings

This study showed that the lowest rate of perinatal death for twins occurred in those delivered between 37 and 38 weeks gestation. Compared with this gestational age at delivery, there was almost a 2-fold increase in perinatal death in deliveries before 37

weeks and a 2-fold increase in perinatal death in twin babies delivered at or beyond 39 weeks after adjusting for potential confounders. Although most of the results were presented in groups of gestational age weeks due to data sparsity, when the results were analysed by individual weeks of gestation, compared to delivery at 37 weeks, the differences in perinatal death were only statistically significant in deliveries before 35 weeks and above 39 weeks gestation. It is likely that the study was underpowered to show a difference in deaths between individual weeks of gestation because of the small number of perinatal deaths in the cohort and particularly in the later gestational weeks.

Guidance from NICE, UK on timing of delivery recommends elective delivery of twins from 37 weeks in dichorionic twins and 36 weeks in monochorionic twins in order to reduce perinatal death (Visintin et al. 2011) and a subsequent systematic review upheld these recommendations (Cheong-See et al. 2016). Our findings agree with delivery at 37-38 weeks to reduce perinatal death as per the national policy but we have not shown that delivery at 36 weeks is significantly different from delivery at 37 weeks in any of the groups. It is important to note however that our sample of monochorionic twins was likely too small to draw conclusions about individual gestational week categories and although overall there was a 2-fold increase in perinatal deaths in monochorionic twins compared to dichorionic twins, from this study we are unable to relate this to gestational age at delivery.

In contrast to some previous studies, we did not find any difference in perinatal death or preterm delivery in twins conceived by ART compared to those spontaneously conceived (Helmerhorst et al. 2004, Suzuki and Miyake 2010, McDonald et al. 2005) and therefore this subgroup of twins should be managed according to the current UK guidelines. Taken together, this information is of use to clinicians planning the antenatal clinical management of twins and/or advising families with twin pregnancy.

6.5.2 Strengths and Limitations

The main strengths of this study are the large, unselected twin sample size in a stable population with high quality data. In particular, data on ART use and indication for induction of labour are rarely available. The retrospective cohort design allowed for efficient use of the routinely collected data. Another strength is the use of accurate gestational age measurements (we excluded pregnancies with inaccurate gestational age measurement) and the completeness of the dataset used. Inaccurate gestational age measurement is often a reason for variations in term and preterm rates between countries (Delnord et al. 2015). Using routinely collected data ensured that every twin pregnancy was included thus reducing the risk of selection bias. We also accounted for the clustering effect of twins within mothers (and hence their similarity to each other) by estimating robust standard errors when producing the estimates and 95% confidence intervals. We believe the results of this study will be generalizable to the UK population.

There are of course some limitations to the use of routinely collected data. Missing covariate values can lead to selection bias if the missing values are not missing at random and can also result in a reduced sample size if included in multivariable analyses leading to a loss of power. In this study, we took the pragmatic approach of not including any covariates with large amounts of missing values, but we did examine the effect of these variables in sensitivity analyses (which corroborated the findings) and we only used records with complete recordings for the stratified analyses. One of the caveats of using routinely collected data is that we are limited in the confounders adjusted for in the model to those that are routinely collected. A potential confounder we were unable to address was place of delivery (a potential confounder as women having a home birth are low risk and therefore different to those delivering in the hospital setting). However, given the small proportion of women who delivered outwith the AMH (99% of deliveries in Grampian are at the AMH which is the source of data collection for the AMND), we believe it is unlikely to have introduced any bias, especially with a twin pregnancy study where a home birth would be very unlikely in any geographical area. Another limitation is the long period of time over which the study population was collected. Obstetric and neonatal care has

likely changed over that time. We adjusted for this in the multivariable analysis by treating year of delivery as a possible confounder.

6.6 Conclusions

In conclusion, perinatal death in twins appears to be lowest in twins delivered from week 37 and by the end of week 38. In keeping with previous studies, perinatal death was 2-fold higher in monochorionic twins compared to dichorionic twins but we did not find any evidence in our study that they should be delivered at differing gestational ages, although the sample size for this subgroup was small. In contrast to some previous studies, we did not find any difference in perinatal mortality between twins born by ART procedures and twins spontaneously conceived and therefore twins conceived by ART should be managed according to the national guidelines. This information should be used when planning antenatal care and counselling women regarding optimum timing of delivery of twin pregnancies.

6.7 Chapter Conclusion

The work presented in **Chapter 6** suggests that, in order to minimise perinatal death, uncomplicated twin pregnancies should not be delivered before 37 weeks or after 39 weeks. Monochorionic twins have a 2-fold higher risk of perinatal death compared to dichorionic twins. ART conceived twins had the same perinatal outcomes as naturally conceived twins in terms of perinatal death and preterm birth and should be managed according to the current UK guidance for the management of twin pregnancies.

Chapter 6 explored the use of routinely collected Scottish maternity data to study short-term perinatal outcomes in twin pregnancies according to gestation at delivery after adjusting for potential confounders. Repeating the exploration of short-term perinatal outcomes from **Chapter 6**, **Chapter 7** used the full Scottish population to repeat the study and provide long-term outcome data by record linking the maternity records to the education records of the twin offspring.

Chapter 7

Gestational Age at Birth of Twins: Perinatal and Childhood Outcomes: a Population Cohort Study of 43,133 Twins

The following materials have been submitted for publication in The Journal of the American Medical Association under the same title by Dr Sarah R Murray (SM), Dr Danny MacKay (DM), Dr Sarah Stock (SS), Professor Jill Pell (JP) and Professor Jane Norman (JN). SM conducted the analysis of the data with input regarding the methods from DM and oversight from JP, SS and JN. SM prepared the first draft of the manuscript under the guidance of JN. All authors provided critical insight for the final draft of the manuscript and approved the final submitted article.

In summary, this work demonstrated that in the absence of a medical complication, twins should not routinely be delivered before 37 weeks gestation. To our knowledge this was the first study to investigate both short- and long-term outcomes according to gestation at delivery of twins. The short-term outcome investigated was perinatal mortality and the long-term outcome was a record of special educational need at school (defined as having one or more of intellectual disabilities, dyslexia, physical or motor impairment, language or speech disorder, autistic spectrum disorder and social, emotional or behavioural difficulties). Compared to remaining in utero, birth at any week from 34-37 weeks was associated with an increased odds of perinatal death and special educational need at school.

This work concluded that in line with the current UK recommendations regarding the antenatal management of twin pregnancy, uncomplicated twin pregnancies should not be delivered before 37 weeks gestation, but limited benefit of prolonging pregnancy thereafter. The short- and long-term risks were shown to be increased before 37 weeks and at 37 weeks the risks of stillbirth and neonatal death were balanced. The risk of

the child having a record of special educational need at school was not increased with birth in week 37 compared to remaining in utero or in the weeks of birth thereafter. This information should be considered by women expecting twins and clinicians managing antenatal care of twin pregnancies especially when making decisions regarding timing of birth.

7.1 Abstract

Importance

Twin pregnancies account for 3% of livebirths but experience substantially more perinatal morbidity and mortality than singletons. Optimising the timing of birth is a key strategy in improving twin pregnancy outcome. Current UK and USA policies are based on observational studies of perinatal mortality and not on longer term effects. The impact of timing of birth on long-term childhood outcome in twins is uncertain.

Objective

To determine the optimal gestation for birth of twin pregnancies by calculating the week of birth associated with the lowest risk of short- and long-term adverse outcomes (perinatal mortality and special educational need at school.

Design: Population-based data-linkage cohort study

Setting: Scotland, United Kingdom

Participants: 43,133 twin infants born from 34 weeks onwards between 1980 and 2015.

Exposures

Gestational age at birth in weeks.

Main Outcomes and Measures

Primary outcomes were extended perinatal mortality and a record of special educational need (one or more of intellectual disabilities, dyslexia, physical or motor impairment, language or speech disorder, autistic spectrum disorder and social, emotional or behavioural difficulties) at school (4-18 years old). To infer the impact

of birth, clinical outcomes of twin infants born at each week of gestation from 34 weeks were compared to twin infants remaining *in utero* thereafter.

Results

Maternity and education records were available for 43,133 and 7,421 twins respectively. Compared to remaining *in utero* birth at any week from 34 to 37 weeks was associated with increased odds of perinatal death (i.e. adjusted [adj.] odds ratio [OR] 1.99, 95% Confidence intervals [CI] 1.53-2.69 at 36 weeks [n=8,056]) and increased risk of special educational need at school (i.e. adj. OR 1.39, 95% CI 1.11-1.74, for birth at 36 weeks compared to 37 weeks). In a competing risks analysis, the risks of stillbirth and neonatal death were balanced at 37 weeks.

Conclusions and Relevance

In the absence of a medical complication, twins should not be routinely delivered before 37 weeks gestation. Our findings will help optimise shared decision making around the timing of twin birth.

7.2 Introduction

The rate of twinning continues to rise, in part due to assisted reproduction technologies, with up to 24% of successful in vitro fertilisation procedures resulting in a multiple pregnancy (Fields et al. 2013, Visintin et al. 2011). Although twin pregnancies account for only 3% of live births, twin infants account for approximately 15% of neonatal and special care baby unit (NNU/SCBU) and neonatal intensive care (NICU) admissions (Harrison and Goodman 2015). These admissions, and the increased cesarean section rate amongst twins and triplets mean that multiple pregnancies are associated with a greater healthcare burden; estimated to be around three times that of a singleton pregnancy (RCOG 2017). In addition to these excess costs and increased morbidity, twin pregnancies are associated with higher perinatal mortality, due to higher rates of both stillbirth (infants born after 24 weeks gestation without signs of life) and greater rates of neonatal death (infants born alive who die within 28 days). Optimising the timing of birth is a key strategy in reducing perinatal mortality and has been highlighted as a research priority (Heazell et al. 2015, Visintin et al. 2011). The optimum timing of birth of twin pregnancies is still uncertain. Randomised controlled trials on timing of birth have not been adequately powered to draw definitive conclusions (Dodd et al. 2012, Suzuki et al. 2000), and therefore observational studies have been used to inform current policy recommendations for twin pregnancy (National Institute for Health and Clinical Excellence in the UK [NICE] (Visintin et al. 2011) and the Society for Maternal and Fetal Medicine [SMFM] in the USA (SMFM 2014)). However, these observational studies have some methodological weaknesses; many have used live births in the week of gestation as the denominator rather than the population at risk (which also includes ongoing pregnancies), they do not adjust for the clustered outcomes of twin pregnancies and many observational studies only consider deaths prior to birth (Kahn et al. 2003, Minakami and Sato 1996). A systematic review of these observational studies vulnerable to bias (n = 9), properly conducted observational studies with a low risk of bias (n = 11) and a smaller number of randomised trials (n = 12) has concluded that

in terms of minimising perinatal death, birth should be considered from 37 weeks in uncomplicated dichorionic pregnancies and 36 weeks in monochorionic pregnancies.

Data from singletons has illustrated the need to consider the impact of timing of birth on long-term neurodevelopmental outcomes (for which performance at schools is a surrogate) as well as short-term outcomes such as perinatal mortality. For example, in uncomplicated singleton pregnancies, perinatal mortality is lowest at 38-39 weeks (Stock et al. 2012, Grobman and Caughey 2019, Hannah et al. 2017), but a record of special educational need at school (includes intellectual disabilities, dyslexia, physical or motor impairment, language or speech disorder, autistic spectrum disorder and social, emotional or behavioural difficulties) is lowest for deliveries at 41 weeks gestation (MacKay et al. 2010). Hence, although perinatal death could be reduced by early delivery in singletons, such a strategy would be associated with an increase in neurodevelopmental compromise for the baby. Current guidelines for singleton pregnancy recommend that routine induction of labour should be deferred until 41 weeks gestation in cases of prolonged pregnancy (NICE 2010). Our objective was to explore the association between gestation at birth of twins and short- and long-term childhood outcomes using a national birth cohort. Such information would be of use to women and their families as well as clinicians and policy makers to guide timing of birth and optimise shared informed decision-making regarding timing of birth of twins.

7.3 Methods

The study was approved by the National Health Service Scotland Public Benefit and Privacy Panel and the South-East Scotland Multi-Centre Research Ethics Committee. All data were nonidentifiable and individual patient consent was not required. A data processing agreement was produced between the University of Edinburgh and Information Services Division (holders of the maternity data) and a data sharing agreement between the University of Edinburgh and Scottish Exchange of Educational Data (ScotXed).

7.3.1 Study Population

A population-based cohort study of twin pregnancies in Scotland, United Kingdom delivered at 34 weeks' gestation or greater between 1 January 1980 and 31 December 2015.

7.3.2 Databases

We obtained data from the Scottish Morbidity Record 02 (SMR02), the Scottish Stillbirth and Infant Death Survey (SSBID) and the Scottish Exchange of Educational Data (ScotXed). The study population was derived from the SMR02 maternity database which collects data on maternal, obstetric and neonatal outcomes. The SSBID database contains information on stillbirths and infant deaths that are registered with the General Register Office for Scotland, with registration mandated by law. The ScotXed database contains details on the pupil census which is conducted annually by all local authority-run primary, secondary and special schools. The information includes whether a child has a special educational need (SEN) and the type. The SMR02 is subjected to regular quality assurance checks and has been more than 99% complete since 1980 (Cole 1980). Using SMR02 as the base population, the education data from the ScotXed database was record-linked to the SMR02 by Information Services Division Scotland. The linkage methodology has been described in detail previously (Wood et al. 2013). The follow up study of the education outcomes was limited to sex discordant twins as for twin infants of the same sex we could not be certain that the correct twin records had been linked.

7.3.3 Inclusion and Exclusion Criteria

Women were included if they had a twin birth at or after 34 weeks gestation. Pregnancies and births complicated by congenital anomaly were excluded. Births were excluded if the gestational age at birth was recorded as missing or greater than 44 weeks, maternal age less than 10 years, parity was missing or listed as greater than

14, birthweight greater than 5000g or if fetal sex was recorded as missing. For the follow up study we excluded individuals whose age was recorded as younger than 4 years or older than 19 years in the pupil census.

7.3.4 Outcomes, Exposures and Covariates

The exposure of interest was gestational age at birth. In the SMR02, the units for gestational age at birth are completed weeks of gestation, in whole units rounded to the nearest week and calculated using the estimated date of delivery recorded in the clinical record. This variable has been described previously and is considered to be accurate and of high quality (MacKay et al. 2010) with more than 95% of women in the United Kingdom having had gestational age confirmed by ultrasound in the first half of pregnancy since the early 1990s (Campbell and Soothill 1993). Gestational age at birth was treated as an ordinal variable. We compared birth at a particular week of gestation to ongoing pregnancies (i.e. women who go on to deliver at a later gestation by any mode of onset, this method has been described previously (Stock et al. 2012, Hannah et al. 2017)).

The two primary outcomes were extended perinatal mortality by week of gestation and a record of special educational need (SEN) at school. Extended perinatal mortality was defined as combined antepartum/intrapartum stillbirth (infant born showing no signs of life) or neonatal death (death of a liveborn infant in the first four weeks of life). SEN is defined by the Department of Education as being unable to benefit fully from school education without help beyond that normally given to schoolchildren of the same age (http://www.teachernet.gov.uk/_doc/3724/SENCodeOfPractice.pdf). In this study SEN was defined as a record of any of the following; intellectual disabilities, dyslexia, physical or motor impairment, language or speech disorder, autistic spectrum disorder and social, emotional or behavioural difficulties.

Secondary outcomes studied included neonatal morbidity (a composite measure of neonatal morbidity defined as an infant with a low Apgar score [<7 at 5 mins] or who

required admission to a neonatal unit [special care or neonatal intensive care unit] or who required assisted ventilation) for each gestational age category at birth. An exploratory analysis of academic achievement at school and school leaver destination (recorded at 6 months after leaving school) of the twin children was also performed. Highest academic attainment was derived from the number of examination awards attained at each level of the Scottish Credit Qualifications Framework and converted into a binary variable: poor educational attainment (consisting of low [≥ 1 at level 2, <5 at level 3, or <2 at level 4] and basic attainment [≥ 5 at level 3, ≥ 2 at level 4, or ≤ 4 at level 5]) or high educational attainment (consisting of broad attainment [>7 at level 4, >5 at level 5, or <3 at level 6] and highest attainment [>1 at level 7 or ≥ 3 at level 6]). School leaver destination of the twin infants according to gestation at birth was defined as a dichotomous variable of higher/further education/employment/training or unemployment.

The following variables were considered to be potential confounders and were included in the multivariate regression analyses: maternal age at birth (≤20, 21-30, 31-40 or >40 years), parity during the index pregnancy (para 0 or para ≥ 1), year of birth (1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, 2006-2010 or 2011-2015), area socioeconomic deprivation quintile of postcode of residence (defined by Carstairs 2001, 1[most affluent] to 5[most deprived] (Morris and Carstairs 1991)), gestation and sex-specific birthweight centiles (<3, 4-10, 11-90, 91-97 or >97), fetal sex (male or female), maternal smoking status at booking (current or non-smoker), maternal height (<150, 150-154, 155-159, 160-164, 165-169, 170-174 or >175cm).

7.3.5 Statistical Analyses

Summary statistics were derived and compared by the outcome of perinatal death using chi squared test for categorical data and chi-squared test for trend for ordinal data. To determine the risk of perinatal death in babies delivered at each gestation compared to ongoing pregnancies univariate and multivariate generalised estimating

equations (GEE) analyses were performed to adjust for the clustering between the twins and presented as odds ratios and 95% confidence intervals. The user-written quasi-likelihood under the independence model criterion (QIC) statistic was used to compare different correlation structures (Cui 2007). The structure with the lowest trace QIC was selected. Covariates included in the multivarible analyses were infant sex, maternal age and height, smoking status, parity, birth-weight centile, year of delivery and deprivation quintile. To assess for the effect of medically indicated deliveries an interaction term was included in the GEE model and compared to the model without the interaction term. The indication for elective caesarean section or induction of labour is not recorded in SMR02, unlike medical conditions in pregnancy (recorded using the international classification of diseases [ICD], ninth and tenth revisions (2010)). Medically-indicated delivery was defined as at least one of the following conditions: hypertensive disease (ICD-10 o10), diabetes mellitus (ICD-10 o24), small for gestational age (ICD-10 p05), thromboembolic disease (ICD-10 o22), liver disorders (ICD-10 o26), antenatal investigation of abnormality (ICD-10 o42) and poor obstetric history (previous stillbirth or neonatal death ICD-10 oo1). A competing risk analysis was then performed assessing the risk of delivery versus expectant management at a particular gestational week. We defined the 'competing risk' of perinatal death at a given gestational week as the difference between the risk of stillbirth and risk of neonatal death for deliveries in that week thus providing a direct measure of benefit or harm from delivery versus expectant management. Univariate and multivariate GEE analyses were performed to determine the relationships between gestational age at birth and a record of SEN at school. School leaver status and academic achievement were analyzed using multivariable logistic regression modelling and presented as odds ratios (OR) and 95% confidence intervals (95% CIs).

Missing values for maternal height, smoking status and deprivation category were created using multiple imputation by chained equations through the use of the ICE module in STATA (Royston 2007). All covariates and outcomes were included in the imputation and 30 imputed datasets created. A sensitivity analysis of complete cases with the imputed datasets was conducted. All available demographic and clinical

variables were used to inform the imputation process. The multivariable models were fitted to each imputed data set, and a pooled result was obtained for estimates of effect.

Chorionicity is not recorded in SMR02, but as chorionicity affects the risk of perinatal mortality (with a two-fold higher perinatal mortality in monochorionic compared to dichorionic twins (Kilby 2017)), a subgroup analysis of only dichorionic twins (identified by sex discordance as they are by definition dizygotic and therefore dichorionic) was performed to determine the association between gestational week at birth and perinatal death in this group.

Population attributable fractions (Brady 1998) were estimated using individuals with complete data to determine what proportion of perinatal death was potentially explained by gestation at birth.

7.3.6 Sensitivity Analyses

To assess the effect of pregnancies complicated by one perinatal death which occurred more than one week before birth (i.e a stillbirth in a different week than the delivery week), a sensitivity analysis was performed to exclude pregnancies with extreme birth weight discordance at birth. Specifically, where there was one fetal death and extreme birth weight discordance (defined as a difference in birthweight of >40% [calculated using the difference in the mean twin birthweight in the Scottish population at 28 and 32 weeks in the cohort which equated to 40% difference]) we assumed the intrauterine fetal death was likely to have occurred more than one week before birth and inclusion of this twin pair would lead to an overestimate in the stillbirth rate in the timeframe studied.

P values for hypothesis tests were two sided and set at $P<0.05$. All analyses were undertaken using STATA MP, version 14.1 (stata corporation).

7.4 Results

The SMR02 contained 43,436 records of twin infants in Scotland in the study period of 1980-2015, 43,133 were eligible for inclusion in the analysis. There were 472 perinatal deaths, 354 stillbirths and 118 neonatal deaths, in this cohort. The process of deriving the study cohort is outlined in Figure 7-1.

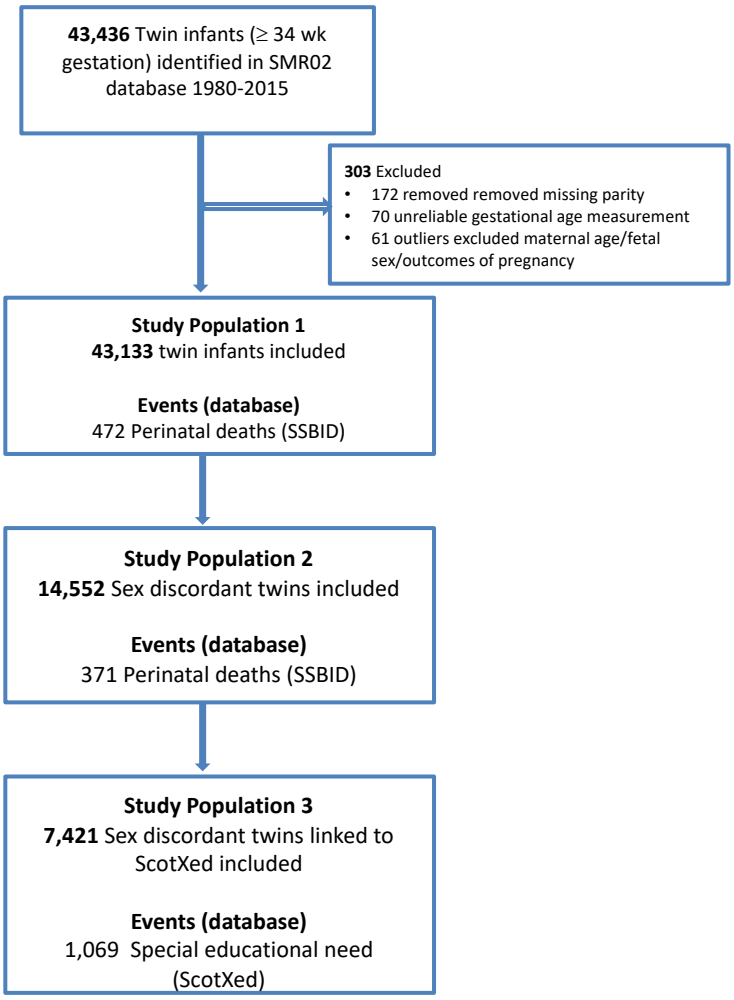


Figure 7-1: Derivation of study cohort: derivation of study populations for the different outcomes. SMR02 indicates Scottish Morbidity Record 02; SSBID, Scottish Stillbirth and Infant Death Survey; ScotXed, Scottish Exchange of Educational Data.

The largest proportion of twins were delivered between 37 and 38 weeks (n = 21,057, 48.8%) and 16,961 (39.3%) were born preterm and 5,115 (11.9%) were born after 38

weeks. There were 26,578 (61.1%) mothers aged between 25 and 35 years of age and 24,576 (57.0%) were primigravid. The characteristics of the cohort are presented in Table 7-1. Data were missing on deprivation category (1.8%), maternal height (14.5%) and maternal smoking (34.3%).

Table 7-1: Baseline demographics of the population of 43,133 twins born in Scotland and the odds of perinatal death

Variable	N (%)	N of perinatal deaths (%)	Crude odds ratio (95% CI)	P Value*
Gestation				
34	3774 (8.75)	85 (2.25)	3.25 (2.38-4.43)	<0.001
35	5131 (11.90)	85 (1.66)	2.37 (1.74-2.34)	
36	8056 (18.68)	103 (1.28)	1.82 (1.36-2.45)	
37	10,925 (25.33)	77 (0.70)	1	
38	10,132 (23.49)	73 (0.72)	1.02 (0.74-1.41)	
39	3261 (7.56)	29 (0.89)	1.26 (0.82-1.94)	
>40	1854 (4.30)	20 (1.08)	1.54 (0.94-2.52)	
Missing	0			
Sex				
Male	21,437 (49.70)	247 (1.15)	1.1 (0.93-1.33)	0.251
Female	21,696 (50.30)	225 (1.04)	1	
Missing	0			
Parity				
Prim	18,287 (42.20)	202 (1.10)	1.02 (0.85-1.22)	0.860
Nulliparous	24,576 (57.60)	270 (1.19)	1	
Missing	0			
Maternal Age				
<20	2235 (5.18)	29 (1.30)	1.27 (0.85-1.90)	0.769
20-25	7610 (17.64)	91 (1.20)	1.17 (0.89-1.52)	
25-30	13336 (30.92)	137 (1.03)	1	
30-35	13242 (20.70)	144 (1.09)	1.06 (0.84-1.43)	
35-40	5915 (13.71)	61 (1.03)	1.01 (0.74-1.36)	
>40	795 (1.84)	10 (1.26)	1.23 (0.85-1.90)	
Missing	0			
Year of birth				
1980-1985	6,140 (14.24)	131 (2.13)	4.31 (2.89-6.42)	<0.001
1986-1990	5,692 (13.20)	73 (1.28)	2.57 (1.68-3.93)	
1991-1995	6,277 (14.55)	73 (1.16)	2.33 (1.52-3.56)	
1996-2000	6,194 (14.36)	65 (1.05)	2.10 (1.36-3.24)	
2001-2005	5,900 (13.68)	42 (0.71)	1.42 (0.89-2.27)	
2006-2010	6,970 (16.16)	58 (0.83)	1.66 (1.07-2.58)	
2011-2015	5,960 (13.82)	30 (0.50)	1	
Missing	0			
Deprivation Carstairs 2001				
1	8568 (19.86)	83 (0.97)	1	0.149
2	7908 (18.33)	94 (1.19)	1.23 (0.91-1.65)	
3	8199 (19.01)	79 (0.96)	1.00 (0.73 – 1.36)	
4	8332 (19.32)	85 (1.02)	1.05 (0.78 – 1.43)	
5	9344 (21.66)	120 (1.28)	1.33 (1.03 – 1.76)	
missing	782 (1.81)			

Birth-weight Centiles				
1-3	1,337 (3.10)	125 (9.35)	22.9 (17.7 – 29.7)	<0.001
4-10	3,078 (7.14)	66 (2.14)	4.9 (3.6 – 6.6)	
11-20	4,360 (10.11)	37 (0.85)	1.9 (1.3 - 2.8)	
21-80	25,900 (60.05)	116 (0.45)	1	
81-90	4,256 (9.87)	20 (0.47)	1.0 (0.7 - 1.69)	
91-97	2,941 (6.82)	14 (0.48)	1.1 (0.6 – 1.9)	
98-100	1,261 (2.92)	94 (7.45)	17.9 (13.6 – 23.7)	
Missing	0			
Smoking status at booking				
Smoker	5912 (13.71)	68 (1.15)	1.54 (1.16-2.05)	0.003
Non-smoker	22416 (51.97)	168 (0.75)	1	
Missing	14805 (34.32)			
Maternal Height				
<150	456 (1.24)	9 (1.97)	1.87 (0.94-3.71)	0.029
150-154	2744 (7.45)	38 (1.38)	1.30 (0.90-1.88)	
155-159	6609 (17.94)	84 (1.27)	1.20 (0.90-1.58)	
160-164	10982 (29.81)	117 (1.07)	1	
165-169	9265 (25.15)	83 (0.90)	0.84 (0.63-1.11)	
170-174	4923 (13.36)	49 (1.00)	0.93 (0.67-1.31)	
>174	1865 (4.32)	13 (0.70)	0.65 (0.37-1.16)	
Missing	6289 (14.58)			

7.4.1 Short Term Perinatal Outcomes According to Gestation at Birth

Outcomes of perinatal mortality by week of gestation compared to ongoing pregnancies are reported in Table 7-2.

Table 7-2: Perinatal mortality, perinatal morbidity (composite of apgar score <7, assisted ventilation or admission to the neonatal unit) and population attributable fraction (PAF) for perinatal mortality at each week of gestation compared to remaining in utero

	Ongoing pregnancies N[a] with outcome/total no in group (%)	Delivered N with outcome/total no in group (%)	OR[b] (95% CIs[c])	P value	Adjusted OR* (95% CIs)	P value	PAF[d]
34	387/39359 (0.98)	85/3774 (2.25)	2.32 (1.80-3.00)	<0.001	2.59 (1.99-3.39)	<0.001	0.101 (0.630-0.138)
35	302/34228 (0.88)	85/5131 (1.66)	1.89 (1.45–2.44)	<0.001	2.12 (1.63-2.76)	<0.001	0.103 (0.543-0.149)
36	199/26172 (0.76)	103/8056 (1.28)	1.69 (1.31-2.18)	<0.001	1.99 (1.53-2.59)	<0.001	0.138 (0.066-0.205)
37	122/15247 (0.80)	77/10925 (0.70)	0.88 (0.65-1.18)	0.397	1.10 (0.81-1.51)	0.543	-0.0523 (-0.175-0.057)
38	49/5115 (0.96)	73/10132 (0.72)	0.75 (0.52-1.09)	0.129	0.92 (0.61-1.38)	0.520	-0.197 (-0.485-0.035)
39	20/1854 (1.08)	29/3261 (0.89)	0.82 (0.45-1.49)	0.520	0.77 (0.41-1.45)	0.412	-0.126 (-0.575-0.195)
Perinatal Morbidity							
34	11712/39359 (30)	3244/3774 (86)	14.45 (12.76-16.36)	<0.001	16.23 (14.23-18.45)	<0.001	0.142 (0.141-0.142)
35	8525/34228 (25)	3187/5131 (62)	4.94 (4.56-5.36)	<0.001	5.67 (5.21-6.17)	<0.001	0.163 (0.162-0.164)
36	5431/26172 (21)	3094/8056 (38)	2.38 (2.22-2.55)	<0.001	2.77 (2.58-2.99)	<0.001	0.167 (0.165-0.169)
37	2948/15247 (19)	2483/10925 (23)	1.23 (1.14-1.33)	<0.001	1.50 (1.38-1.63)	<0.001	0.068 (0.065-0.071)
38	1087/6976 (16)	1861/8271 (23)	0.83 (0.75-0.93)	0.001	1.01 (0.90-1.14)	0.824	-0.099 (-0.107- -0.908)
39	398/1404 (28)	689/2572 (27)	0.98 (0.82-1.17)	0.825	1.05 (0.87-1.27)	0.599	-0.010 (-0.023-0.002)

[a]N= number, [b]OR = odds ratio, [c]CI = 95% confidence intervals, [d]PAF = population attributable fraction.

*Adjusted for maternal age & height, parity, baby sex, birth order, year of delivery, deprivation Category, birth weight centiles, smoking

Primary Outcome

Compared to remaining *in utero*, birth of twin infants at any complete week from 34 to 37 weeks gestation was associated with an increased risk of perinatal mortality (adjusted odds ratio [adj. OR] 2.59, 95% CI 1.99 – 3.39 at 34 weeks [n=3,774], adj. OR 2.12, 95% CI 1.63-2.76 at 35 weeks [n=5,131] and adj. OR 1.99, 95% CI 1.53-2.69 at 36 weeks [n=8,056]). Birth at 37, 38 or 39 weeks gestation was associated with no increased risk of perinatal mortality compared to remaining *in utero* (adj. OR 1.10, 95% CI 0.81-1.51 at 37 weeks, adj. OR 0.92, 95% CI 0.61-1.38 at 38 weeks and adj. OR 0.77, 95% CI 0.41-1.45 at 39 weeks). Gestation at birth before 37 weeks had a population attributable fraction (PAF) of perinatal mortality of 34.2% (Table 7-2).

In the competing risk analysis, the risk of stillbirth was significantly lower than the risk of neonatal death at 34 and 35 weeks gestation (risk difference -2.49 at 34 weeks and -2.43 at 35 weeks, table 4) but balanced at 37 weeks (risk difference 2.05, 95% CI 0.8-3.3). After 37 weeks the risk of stillbirth (if birth does not occur) significantly outweighed the risk of neonatal death following birth (Figure 7-2).

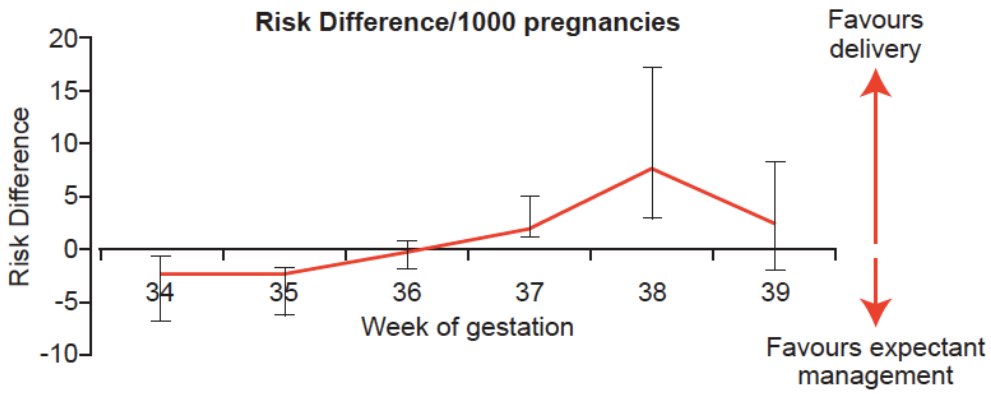


Figure 7-2: Competing risk analysis results

Secondary Outcomes

Table 7-2 reports neonatal morbidity for babies delivered and babies remaining *in utero* for each week of gestation, together with the effect of birth on the odds of neonatal morbidity. Twin infants born before 38 weeks had an increased risk of neonatal morbidity compared to remaining *in utero* (adj. OR 16.23, 95% CI 14.23 – 18.45 at 34 weeks, adj. OR 5.67, 95% CI 5.21-6.17 at 35 weeks, adj. OR 2.77, 95% CI 2.58-2.99 at 36 weeks and adj. OR 1.50, 95% CI 1.38-1.63 at 37 weeks. Birth at 38 and 39 weeks had no increased or decreased risk of neonatal morbidity compared to remaining *in utero* (adj. OR 1.01, 95% CI 0.90-1.14 at 38 weeks and adj. OR 1.05, 95% CI 0.87-1.27 at 39 weeks).

7.4.2 Sex Discordant Twins

Results of the subgroup analysis of sex-discordant (and therefore dichorionic) twins (n = 14,755) were similar to the results for all twins: birth before 37 weeks was associated with an increased risk of perinatal death compared to remaining *in utero* (adj. OR 1.81, 95% CI 1.28-2.55, Appendix Table 7-1). There was no statistically significant difference in perinatal death with birth at 37-38 weeks compared to remaining *in utero* (adj. OR 0.64, 95% CI 0.34-1.12).

7.4.3 Long Term Educational Outcomes According to Gestation at Birth

Primary Outcome

7,421 twin children were linked to their educational data and 1,069 had a record of SEN (14.4%). There was an inverse linear relationship between gestation at birth and SEN (Figure 7-3). Compared to birth at 37 weeks, children born at each week of gestation before 37 weeks had an increased risk of SEN at school (adj. OR 1.35, 95% CI 1.01-1.82 at 34 weeks, adj. OR 1.35, 95% CI 1.05-1.74 at 35 weeks and adj. OR 1.39, 95% CI 1.11-1.74 at 36 weeks, Table 7-3). The risk of SEN did not change with birth between 38 and 39 weeks (Table 7-3) or with births >40 weeks (adj. OR 1.16, 95% CI 0.69-1.96). The overall rate of SEN and the rate at each gestational week was

higher in twins than has previously been reported in singletons (MacKay et al. 2010) (Appendix Table 7-2).

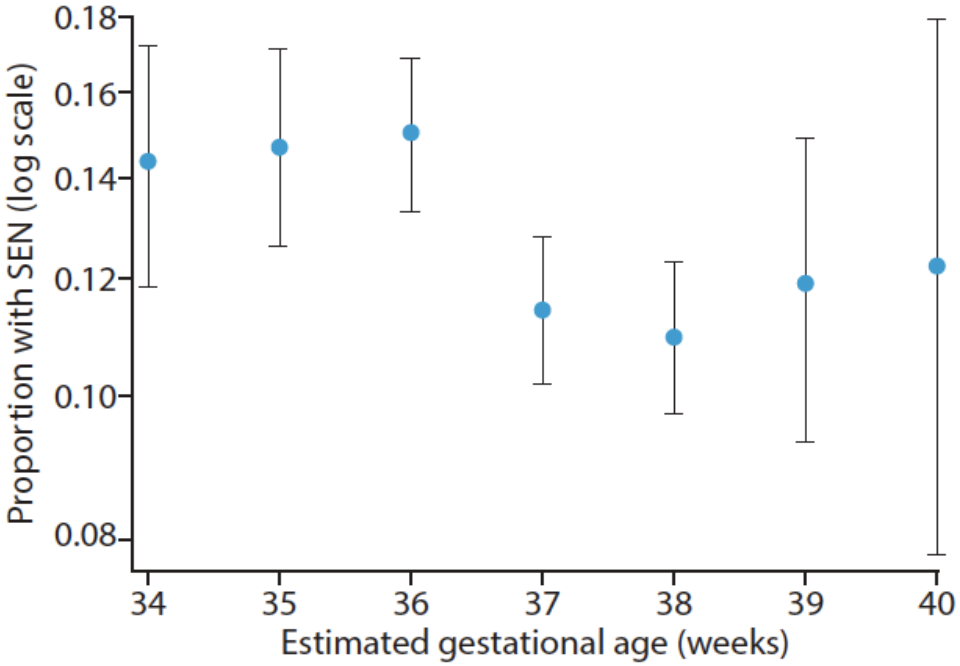


Figure 7-3: Prevalence of special educational need by gestation at birth

Table 7-3: Association between gestation at birth and Special Educational Need (SEN), Leaver status and Academic Attainment

Gestation at Birth (weeks)	N[a] Outcome/Total N (%)	OR[b] (95% CI[c])	Adjusted* OR (95% CI)
SEN			
34	98/682 (14.37)	1.30 (0.97-1.74)	1.35 (1.01-1.82)
35	148/1005 (14.73)	1.34 (1.05-1.71)	1.35 (1.05-1.74)
36	235/1560 (15.06)	1.37 (1.10-1.71)	1.39 (1.11-1.74)
37	263/2301 (11.43)	1	1
38	239/2184 (10.94)	0.95 (0.77-1.18)	1.00 (0.80-1.24)
39	64/538 (11.90)	1.05 (0.75-1.47)	1.02 (0.73-1.44)
>40	22/180 (12.22)	1.08 (0.64-1.82)	1.16 (0.69-1.96)
Leaver Status (not in employment at 6 months)			
34			
35	19/219 (8.68)	0.74 (0.41-1.34)	0.77 (0.42-1.43)
36	22/259 (7.83)	0.66 (0.38-1.15)	0.67 (0.38-1.19)
37	43/385 (10.05)	0.87 (0.56-1.36)	0.91 (0.59-1.43)
38	71/553 (11.38)	1	1
38	51/579 (8.10)	0.69 (0.45-1.04)	0.77 (0.51-1.18)
>40	34/185 (15.53)	1.43 (0.87-2.36)	1.75 (1.06-2.87)
	16/108 (12.90)	1.15 (0.60-2.21)	1.26 (0.65-2.49)
Poorer Academic Attainment			
34	19/223 (8.52)	1.14 (0.60-2.15)	1.07 (0.57-2.02)
35	28/290 (9.66)	0.99 (0.59-1.68)	0.95 (0.54-1.65)
36	30/426 (6.58)	1.51 (0.88-2.58)	1.48 (0.85-2.58)
37	62/646 (9.60)	1	1
38	48/616 (7.79)	1.26 (0.79-2.00)	1.17 (0.73-1.90)
39	20/219 (9.13)	1.06 (0.59-1.90)	0.91 (0.50-1.64)
>40	7/80 (8.75)	1.11 (0.44-2.77)	1.07 (0.43-2.70

[a]N= number, [b]OR = odds ratio, [c]CI = 95% confidence intervals, *Adjusted for maternal age & height, parity, baby sex, birth order, year of delivery, social class, birth weight centiles, smoking

Secondary outcomes

The exploratory analysis of academic attainment contained 2,551 twin children. In this sample, gestational age at birth was not associated with low educational attainment at school (adj. OR 1.07, 95% CI 0.57-2.02 at 34 weeks compared to birth at 37 weeks, Table 7-3). The analysis of school leaver destination contained 2,531 twin children. Being born before 37 weeks was not associated with unemployment (adj. OR 0.77, 95% CI 0.42-1.43 at 34 weeks compared to birth at 37 weeks). Being born at 39 weeks was associated with an increased risk of unemployment compared to birth at 37 weeks (adj. OR 1.75, 95% CI 1.06-2.87).

7.4.4 Sensitivity Analyses

The results were very similar when the logistic regression models were run without imputation of missing values (Appendix Table 7-3).

There was no evidence of interaction of medically indicated deliveries compared to all twin deliveries and therefore a full subgroup analysis was not required (Likelihood ratio test of interaction p = 0.751 at 34 weeks, Appendix Table 7-4).

The results were again similar when we ran the logistic regression models excluding those twin pairs with one perinatal death and extreme birth weight discordance (Appendix Table 7-4).

7.5 Discussion

7.5.1 Summary of the Main Findings

In this population-based cohort study of 43,133 twin infants, birth before 37 weeks gestation was associated with an increased risk of perinatal mortality and an increased risk of the children having special educational need at school. The risks of stillbirth

and neonatal death were balanced at 37 weeks gestation. There was no change in the risk of SEN associated with birth at or after 37 weeks of gestation. There was a statistically significantly increased risk of neonatal morbidity associated with birth at 37 weeks compared to birth after this gestation, but this did not result in any difference in longer term school outcomes of the twin children.

7.5.2 Interpretation

The findings of this study suggest that birth at 37 weeks gestation is associated with optimal short- and long-term outcomes for babies born as twins. These data are in keeping with current clinical guidelines (Visintin et al. 2011, Gynecologists and Medicine 2014) which are informed by short term outcomes, but show for the first time that such a policy optimises long term outcomes too. There is clear evidence that birth of twins before 37 weeks is associated with an increased risk of perinatal mortality and the twin child having a record of SEN at school. In contrast to singletons, where the risk of SEN continues to reduce across all gestational ages until 41 weeks then rises again at 42 weeks (MacKay et al. 2010) in twins SEN is lowest at 37 weeks with no statistically significant increase in SEN beyond that week of gestation. This may suggest that in twins (in contrast to singletons) there is no benefit in prolonging pregnancy beyond 37 weeks. We recognise that our sample size is limited in the later gestational weeks, and caution needs to be exercised in drawing firm conclusions about this.

By estimating how often twin pregnancies are complicated by one twin death which occurred before the week of birth (which would hypothetically lead to an overestimation of term stillbirth rates) we have shown that this is not a common problem and is unlikely to affect the results of previous studies which have not accounted for it.

7.5.3 Findings in the Context of Existing Literature

The result of the competing risk analysis (stillbirth and neonatal death balanced at 37 weeks) is consistent with the effect reported by other studies (Cheong-See et al. 2016). However, these studies lacked information on adjustment for clustered outcomes of twins (this is important as failure to account for this will lead to an overestimate of the effect and confidence intervals that are too narrow for the estimate) and a proportion of the data was from randomised controlled trials, and therefore different from population studies (because it is a sample rather than a whole population and therefore is subject to selection bias which can lead to error and affect generalizability of the study results). A randomised control trial on timing of twin birth is unlikely to be feasible given the large sample size that would be required (Dodd et al. 2012) and therefore population studies such as this will remain the mainstay of investigation. Some studies have recommended birth of certain twin groups from 34 weeks gestation, our study does not provide any evidence of benefit of this policy (Spong et al. 2011, SMFM 2014).

To our knowledge this is the first nationwide study to look at long term educational outcomes of twins across the range of gestational age categories. Overall the population rate of SEN is higher in twins (14.41%) than has been previously estimated in singletons (4.90%)(MacKay et al. 2010) and this rate is higher in twins at each week of gestation (Table e2). The overall twin SEN rate in twins is in fact likely to be higher than what is estimated in this study because we only looked at deliveries from 34 weeks and deliveries between 24-34 weeks would likely be associated with increased rates of SEN as found in singletons.

7.5.4 Strengths and Limitations

This study has a number of strengths. Using routinely collected population data ensured that every twin birth that met the inclusion criteria was included thus reducing the risk of selection bias. Both the obstetric and education data was derived from routine population data sources reducing the risk of recall or ascertainment bias. There were a number of methodological analysis benefits of our study over previous twin studies. Firstly, we were able to account for the clustering effect of twins by

using GEE analyses. Furthermore, by using adjusted sex and gestation specific birth weight centiles rather than absolute birthweight we could adjust for the known interaction between birth weight, gestational age and intellectual impairment (Eide et al. 2007, Stephens and Vohr 2009). Finally by carrying out a sensitivity analysis, we have investigated and estimated how often twin deliveries are complicated by one fetal death at a much earlier gestation than is recorded, most studies fail to account for this (Smith 2005). The inclusion of educational outcomes for twins is novel and provides long term outcome data which is crucial to consider when planning timing of birth. Scotland has a stable population with high quality routine data and there was a very high match rate of education records to health records.

The study has limitations due to the nature of the routinely collected data. Missing covariates are acknowledged as being a limitation and potential loss to study power. Multiple imputation was used in this study to counteract this problem and complete case note analysis undertaken as a sensitivity analysis. Residual confounding is a potential limitation with observational data and we can only adjust for potential confounders recorded in the dataset. Two potential confounders we were unable to address in the study were chorionicity and conception status (whether the twins were conceived by assisted reproduction technologies [ART]). We performed a sex discordant analysis to represent dichorionic pregnancies and found very similar results to all twins (increased perinatal mortality with deliveries before 37 weeks gestation). We were unable to adjust for those born following ART, but recent studies have suggested there is no difference in perinatal outcomes between twins conceived naturally and those conceived by ART (Suzuki and Miyake 2010, Helmerhorst et al. 2004). Another limitation is the long period of time over which the cohort was derived and the potential changes in clinical practice over that period in terms of both obstetric and neonatal care. We adjusted for this in the multivariable model by categorizing year of delivery and including it as a potential confounder. Medical indication for delivery is not recorded in SMR02 and could therefore be a source of residual confounding, we attempted to account for this by identifying ICD-10 codes for medical complications in pregnancy and adjusted for these accordingly. Our sample

size in some of the analyses is a limitation, especially in the group after 38 weeks and this is likely due to the policy of planned delivery around 37-38 weeks. For the SEN analysis we were limited by data linkage and could only use sex discordant twins for this analysis thus reducing our sample size further especially in the weeks beyond 38 weeks where we did not see any significant differences compared to birth at 37 weeks.

7.5.5 Clinical and Research Implications

Our findings suggest that in terms of short- and long-term outcomes uncomplicated twin pregnancies should not be delivered before 37 weeks gestation, but that there is limited benefit of prolonging pregnancy thereafter. Being born before 37 weeks is associated with an increased risk of both perinatal mortality and having a record of SEN at school. The risk of stillbirth and neonatal death are balanced at birth at 37 weeks and the risk of SEN does not increase after birth at 37 weeks gestation. This information should be considered women expecting twins and their care givers when making decisions regarding timing of birth.

There are a number of research priorities for twins which require further investigation. As demonstrated in this study, preterm birth remains a significant problem for twin pregnancies (39.2% of deliveries in this study were between 34 and 37 weeks gestation) and at present there is no evidence of benefit of any interventions to reduce preterm birth in twins. Birth order was included as a potential confounder given the established increased risk of perinatal death for second twins (Smith et al. 2002) but it would be important in the future to consider long term school outcomes in twins according to birth order at birth. Randomised controlled trials of timing of birth are likely to be unfeasible therefore population databases should consider including a record of chorionicity in their data collection to inform meta-analyses going forward (chorionicity is now routinely confirmed by ultrasound for twins pregnancies in accordance with national guidelines (Kilby 2017, Gynecologists and Medicine 2014).

7.6 Conclusions

Twin children born in Scotland between 1980 and 2015 had increased rates of perinatal morality and SEN at school if delivered before 37 weeks gestation. After 37 weeks gestation the risks of stillbirth and neonatal are balanced and there is no increased risk of SEN in deliveries beyond 37 weeks. The optimal gestation for birth of uncomplicated twin pregnancies is 37 weeks.

7.7 Appendices

Appendix Table 7-1. Perinatal mortality at each gestation category compared to remaining *in utero* in dichorionic twins.

	Ongoing pregnancies N[a] with outcome/total no in group (%)	Delivered N with outcome/total no in group (%)	OR[b] (95% CI[c])	P value	Adjusted* OR (95% CI)	P value
34-36	294/34698 (0.98)	77/3527 (2.18)	1.75 (1.26-2.44)	<0.001	1.87 (1.32-2.64)	<0.001
37-38	222/29952 (0.88)	72/4746 (1.52)	0.48 (0.29-0.80)	0.005	0.59 (0.34-1.01)	0.055

[a]N= number, [b]OR = odds ratio, [c]CI = 95% confidence intervals

*Adj. for maternal age & height, parity, baby sex, birth order, year of delivery, social class, birth weight centiles, smoking, height

Appendix Table 7-2. Differences in the proportion of special educational need (SEN) between twins and singletons overall and according to gestation at birth, singleton data taken from MacKay *et al.*(MacKay et al. 2010).

Week of gestation	Proportion SEN twins N (%)	Proportion SEN singletons N (%)
33-36	481/3247 (14.81)	1281/16754 (7.65)
37	263/2301 (11.43)	1217/18617 (6.54)
38	239/2184 (10.94)	2759/48810 (5.65)
39	64/538 (11.90)	3848/77217 (4.98)
>40	22/180 (12.22)	10133/222943 (4.55)
Overall rate	1069/7421 (14.41)	17784/362688 (4.90)

Appendix Table 7-3. Sensitivity Analysis: Complete case note analysis (n = 23,762)

Week of Gestation	Ongoing pregnancies N[a] with outcome/total no in group (%)	Delivered N with outcome/total no in group (%)	OR[b] (95% CI[c])	P value	Adjusted* OR (95% CI)	P value
34	387/39359 (0.98)	85/3774 (2.25)	2.32 (1.80-3.00)	<0.001	2.36 (1.56-3.58)	<0.001
35	302/34228 (0.88)	85/5131 (1.66)	1.89 (1.45–2.44)	<0.001	2.40 (1.65-3.50)	<0.001
36	199/26172 (0.76)	103/8056 (1.28)	1.69 (1.31-2.18)	<0.001	2.39 (1.58-3.59)	<0.001
37	122/15247 (0.80)	77/10925 (0.70)	0.88 (0.65-1.18)	0.397	1.60 (0.93-2.75)	0.087
38	49/5115 (0.96)	73/10132 (0.72)	0.75 (0.52-1.09)	0.129	1.08 (0.43-2.73)	0.868
39	20/1854 (1.08)	29/3261 (0.89)	0.82 (0.45-1.49)	0.520	5.43 (0.15-190.1)	0.315

[a]N= number, [b]OR = odds ratio, [c]CI = 95% confidence intervals *Adj. for maternal age & height, parity, baby sex, birth order, year of delivery, social class, birth weight centiles, smoking, height

Appendix Table 7-4. Subgroup analysis of Non-Medically indicated deliveries (n=38,225 for non-medically indicated deliveries)

Week of gestation	All Twins OR (95% CI), p value	Non-Medically indicated OR (95% CI), p value	P value Interaction
34	2.08 (1.42-3.06), <0.001	2.25 (1.66-3.05), <0.001	0.751
35	1.66 (1.12-2.47), 0.011	1.82 (1.34-2.48), <0.001	0.725
36	1.62 (1.08-2.41), 0.019	1.58 (1.17-2.13), 0.003	0.931
37	0.73 (0.44-1.21), 0.225	0.89 (0.63-1.26), 0.526	0.517
38	0.60 (0.31-1.17), 0.132	0.76 (0.49-1.18), 0.228	0.558
39	0.88 (0.30-2.63), 0.822	0.74 (0.37-1.45), 0.381	0.784

Appendix 7-5. Sensitivity analysis: removal of cases complicated by one perinatal death and extreme birth weight discordance (n=382 removed)

Week of gestation	All Twins Adjusted OR (95% CI)	BW discordant deaths removed Adjusted OR (95% CI)
34	2.59 (1.99-3.39)	2.39 (1.63-3.49)
35	2.12 (1.63-2.76)	1.82 (1.24-2.68)
36	1.99 (1.53-2.59)	1.71 (1.24-2.35)
37	1.10 (0.81-1.51)	1.16 (0.76-1.75)
38	0.92 (0.61-1.38)	1.01 (0.60-1.71)
39	0.77 (0.41-1.45)	0.86 (0.36-2.07)

7.7 Chapter Conclusion

The work presented in **Chapter 7** suggests that uncomplicated twin pregnancies should not be delivered before 37 weeks. Birth of twins before 37 weeks was associated with an increased risk of perinatal death and an increased risk of the children having a record of special educational need at school. The risk of stillbirth and neonatal death in twins were balanced at 37 weeks.

Chapter 7 explored the use of routinely collected population maternity data to study both short- and long-term childhood outcomes in twin pregnancies according to gestation at birth after adjusting for potential confounders. **Chapter 8** used the same full Scottish population database to study perinatal outcomes in twins compared to singletons.

Chapter 8

Perinatal Outcomes in Twins Compared to Singletons According to Gestation at Delivery: a Population Cohort Study of 2,004,587 Infants in Scotland

The following materials are in preparation for submission to PLOS Medicine under the same title by Dr Sarah R Murray (SM), Dr Danny MacKay (DM), Dr Sarah Stock (SS), Professor Jill Pell (JP) and Professor Jane Norman (JN). SM conducted the analysis of the data with input from DM and oversight from JP, SS and JN. SM prepared the first draft of the manuscript under the guidance of JN with JP and SS commenting on the completed draft. All authors provided critical insight for the final draft of the manuscript included here and will approve the final submitted article.

In summary, this work demonstrated that twins have higher rates of stillbirth and neonatal death overall compared to singletons. At every week of gestation studied, the stillbirth risk in twins was higher than in singletons born at the same gestation. Neonatal death however was higher in twins compared to singletons for birth at extreme preterm gestation and term gestation (>37 weeks) but lower in twins compared to singletons for births between 29-37 weeks.

This work concluded that in contrast to previous studies, but in line with what would be expected clinically, the odds of stillbirth in twins were higher at every week of gestation compared to singletons. We speculate the lower rates of neonatal death in twins compared to singletons observed between 29-37 weeks reflects different contributing causes of the preterm birth in each group.

8.1 Abstract

Background

The aim of this study was to examine the risk of stillbirth and neonatal death (NND) in twins compared to singletons across the full range of gestation (term and preterm). Although twins at term have higher rates of perinatal death than singletons, previous studies have suggested that the risk might be reversed for preterm birth.

Methods and Findings

We conducted a population-based retrospective cohort study using routinely collected Scottish maternity data. The odds of stillbirth and NND in 52,296 twins were compared to 1,952,291 singletons across the range of gestation from 24 weeks to 42 weeks. Generalised estimating equation (GEE) multivariate logistic regression analysis was performed. Overall twins had higher odds of stillbirth (odds ratio [OR] 3.28, 95% confidence interval [CI] 3.04-3.54 and NND (OR 6.87, 95% CI 6.37-7.40) compared to singletons. Twins had greater odds of stillbirth at every week of gestation compared to singletons, most marked for early term delivery (37-38 weeks) with an adjusted (adj.) OR of 18.06 (95% CI 14.26-22.88). The odds of NND were higher in twins compared with singletons for extreme births (<28 weeks, adj. OR 1.39, 95% CI 1.20-1.61) and delivery at 39-42 weeks gestation (adj. OR 3.00, 95% CI 1.73-5.18). In contrast, the odds of NND were lower in twins compared to singletons for deliveries at 29-37 weeks gestation.

Conclusions

Overall twins have higher rates of stillbirth and NND compared to singletons. Contrary to previous studies, we found that for each gestation studied, the stillbirth risk in twins was higher than in singletons born at the same gestation. In contrast, the odds of NND were higher in twins compared to singletons for birth at extreme preterm

gestation (<28 weeks) and at 39-42 weeks gestation but lower in twins compared to singletons from 29-37 weeks. We speculate this is due to the different aetiology of preterm birth in twins compared to singletons. An alternative, but less plausible, hypothesis is that there are different approaches to neonatal care between twins and singletons specific to preterm birth.

8.2 Introduction

In the UK, the latest MBRRACE report (Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK) stated that the perinatal mortality rate in twins was three times higher than singletons (Manktelow et al. 2014). Much of this higher perinatal mortality is thought to be driven by prematurity with 50% of twins delivering preterm (before 37 weeks of gestation)(ISD Scotland 2009) compared to 6% in singleton pregnancies in the UK (Chang et al. 2013). Despite the established higher risk of perinatal death and preterm birth in twins compared to singletons, few studies have investigated the differences in perinatal mortality between twins and singletons by gestation at birth. This lack of information about twin-specific risk of stillbirth and/or neonatal death (NND) compromises informed clinical decision making in the management of twin pregnancy, and prevents accurate and effective counselling of parents with twins.

Perinatal mortality is a rare event and observational data is often used to study this outcome (Smith 2005). When investigating perinatal mortality, it is essential to ensure the correct denominator, or population at risk, is used in the analysis. It is also key to stratify by gestational age at delivery given that prematurity, especially prevalent in twins, is the largest cause of NND. Studies failing to stratify by gestation assume the same relative risk among twins at each gestational age. Previous studies have found a lower perinatal mortality rate in twins born preterm (<37 weeks) compared to singletons born preterm (Vasak et al. 2017, Kahn et al. 2003, Minakami and Sato 1996). This is a surprising result given it is widely accepted that twins have a greater overall risk of perinatal mortality. However, we believe these studies have some methodological weaknesses because they used live births in the week of gestation as the denominator for perinatal death (rather than the population at risk which also includes ongoing pregnancies), they do not adjust for the clustered outcomes of twin pregnancies (failing to do so can lead to an over estimate of the

effect) and most studies combined stillbirth and neonatal deaths together in the numerator in spite of them having different denominators.

The aim of our study was to assess the perinatal outcomes of stillbirth and NND in twins compared to singletons across the whole range of gestation at birth; including both term and preterm gestations.

8.3 Methods

8.3.1 Data Sources

We carried out a population-based cohort study using routinely collected Scottish Maternity data and included twin and singleton babies born in Scotland between 1st January 1980 and 31st December 2015. Data were obtained from the Scottish Morbidity Record 02 (SMR02) and the Scottish Stillbirth and Infant Death Survey (SSBID). The study population was derived from the SMR02 maternity database which collects data on all maternity admissions to Scottish hospitals (therefore providing maternal, obstetric and neonatal outcomes). The SMR02 covers 98-99% of all births in Scotland (the remainder being home births that do not generate a hospital admission). The database is subjected to regular quality assurance checks and has been more than 99% complete since 1980 (Cole 1980). The SSBID database contains information on stillbirths and infant deaths that are registered with the General Register Office for Scotland, with registration mandated by law. The inclusion criteria consisted of twin and singleton infants born between 24 and 43 weeks' gestation. Pregnancies and births complicated by fetal anomaly were excluded. Births were excluded if the gestational age at delivery was recorded as missing, maternal age was recorded as less than 10 years, parity was missing or recorded as greater than 14, birthweight recorded as greater than 5000g or fetal sex was recorded as missing. In the SMR02 database gestational age at delivery is defined as completed weeks of gestation (with 24 weeks gestation meaning 24 weeks + 0 days to 24 weeks +6 days) on the basis of the estimated date of delivery recorded in each women's

clinical record. This variable has been described previously and is considered to be of high quality (MacKay et al. 2010). The definition of preterm was delivery less than 37 weeks' gestation. The data extract was anonymised prior to being provided to the investigators and therefore individual patient consent was not required. Permission to access the data was granted by the National Health Service Scotland National Public and Privacy Panel and the South-East Scotland Multi-Research Ethics Committee (NHS REC ref 15/SS/0197).

The main exposure of interest was twin pregnancy. The analyses were carried out in all deliveries, in gestational age categories (extreme preterm [<28 weeks], very preterm [29-32 weeks], moderate preterm [33-34 weeks], late preterm [35-36 weeks], early term [37-39 weeks] and full term [39-43 weeks]) and then separately at each week of gestation to determine the perinatal mortality differences between twins and singletons born at different gestations. The two primary outcomes were stillbirth (combined antepartum/intrapartum stillbirth [infant born showing no signs of life]) and extended NND (death of a liveborn infant in the first 4 weeks of life). For stillbirths, we compared the proportion of stillbirths in a given week of gestation to pregnancies at risk of stillbirth within that week in twins compared to singletons. We chose not to use livebirths to calculate the stillbirth rate as the population at risk of antenatal stillbirth is all ongoing pregnancies at any week of gestation as opposed to the fraction of babies who were delivered. This methodological approach is now widely accepted (Kramer et al. 2002). For NND, we used the denominator of live births within that week and again compared twins to singletons.

The following variables were considered, a priori, to be potential confounders and were included as covariates in the multivariate regression analyses: maternal age at delivery (≤20, 21-30, 31-40 or >40 years), parity during the index pregnancy (para 0 or para ≥ 1), year of birth (1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, 2006-2010 or 2011-2015), area socioeconomic deprivation quintile of postcode of residence (defined by Carstairs 2001, 1[most affluent] to 5[most deprived] (Morris and Carstairs 1991)), gestation and sex-specific birthweight centiles (<3, 4-10, 11-90,

91-97 or >97), fetal sex (male or female), smoking status at booking (current smoker or non-smoker), maternal height (<150, 150-154, 155-159, 160-164, 165-169, 170-174 or >175cm).

8.3.2 Statistical Analyses

Summary statistics were derived and compared by twin and singleton pregnancies using chi squared tests for categorical data and chi squared tests for trend for ordinal data. To evaluate the association between twin pregnancy and stillbirth and NND univariate and multivariable generalised estimating equation (GEE) analyses were performed and the results were presented as odds ratios (OR) and 95% confidence intervals (CIs). A GEE model was chosen, using the mother ID as the clustering variable to account for the non-independence of the twin infants. The user-written quasi-likelihood under the independence model criterion (QIC) statistic was used to compare different correlation structures (Cui 2007). The structure with the lowest trace QIC was selected. Covariates included in the multivariate analyses were infant sex, maternal age and height, parity, birth-weight centile, year of delivery and socioeconomic deprivation quintile. To assess for the effect of medically indicated deliveries an interaction term was included in the GEE model and compared to the model without the interaction term. It is known that twins have higher rates of iatrogenic preterm birth (up to one third of preterm deliveries in twins are iatrogenic) compared to singletons; hence we decided to exclude iatrogenic preterm birth in a subgroup analysis (Fuchs and Senat 2016). The indication for elective caesarean section or induction of labour is not recorded in SMR02, unlike medical conditions in pregnancy (recorded using the international classification of diseases [ICD], ninth and tenth revisions (2010)). Therefore, record of the following conditions was assumed to confer a medical indication for delivery: hypertensive disease (ICD-10 o10), diabetes mellitus (ICD-10 o24), small for gestational age (ICD-10 p05), thromboembolic disease (ICD-10 o22), liver disorders (ICD-10 o26), antenatal investigation of abnormality (ICD-10 o42) and poor obstetric history (previous stillbirth or neonatal death ICD-10 oo1).

Records with missing data were included in the summary statistics but excluded from the univariate and multivariate analyses. Maternal smoking was not included as a covariate in the final model due to the large number of missing values but a sensitivity analysis with and without including maternal smoking as a covariate was performed.

Chorionicity is not recorded in SMR02, but as chorionicity affects risk of perinatal mortality, subgroup analysis of only dichorionic twins (identified by sex discordance as they are by definition dizygotic and therefore dichorionic) was performed to determine the association between stillbirth and NND in dichorionic twins compared to singletons.

P values for hypothesis tests were two sided and statistical significance defined as $P<0.05$. All analyses were undertaken using STATA MP, version 14.1 (stata corporation).

8.4 Results

The Study population consisted of 2,011,303 infants of which 2,004,587 were eligible for inclusion in the analysis (52,296 twins and 1,952,291 singletons) (Figure 8-1). There were 9,164 stillbirths and 5,317 NNDs in the cohort. In singleton pregnancies, there were 12,930 perinatal deaths (perinatal mortality rate of 6.7/1000) compared to 1,551 perinatal deaths in twins (perinatal morality rate of 29.7/1000, Appendix Table 8-1).

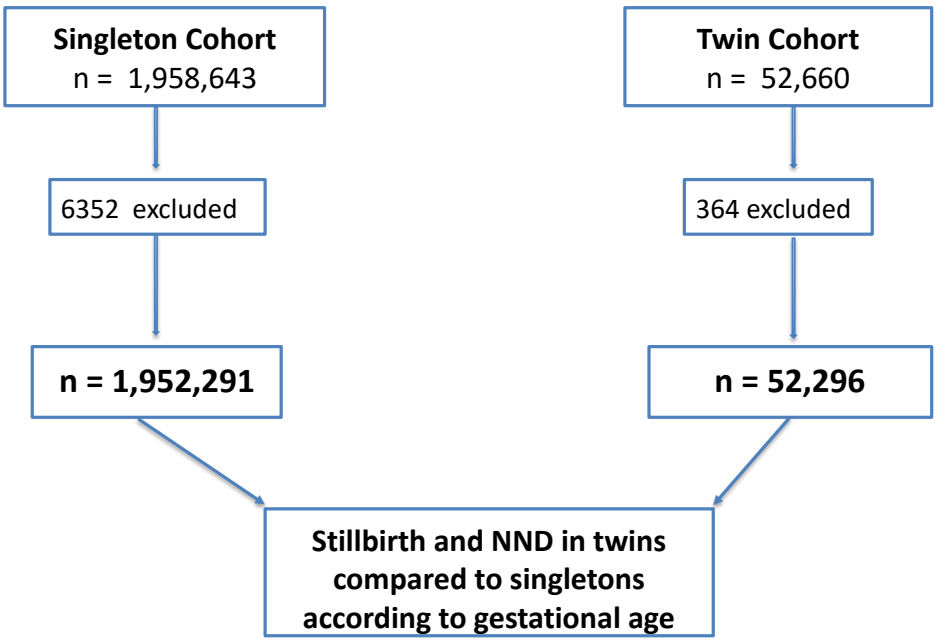


Figure 8-1: Derivation of the Study Cohort

Table 8-1 summarises the pregnancy characteristics of the cohort. Among the twin infants, the largest proportion of infants were born between 37-38 weeks (40.3%) compared to >40 weeks in singletons (77.4%). Of all twins, 49.9% were born preterm (<37 weeks of gestation) compared to 5.9% of singletons. Twin mothers were older than mothers of singleton babies (31.9% versus 22.5% aged 30-34 respectively, Table 1). Data were missing on deprivation category (1.8%), maternal smoking (52.1%) and birthweight (0.1%).

Table 8-1: Baseline summary statistics of the population of 2,004,587 singleton and twin infants born in Scotland

Pregnancy Characteristic	Total N[b]	N (%) in each group		P value
		Singletons	Twins	
Stillbirth				
No	1 995 423	1 943 861 (99.6)	51 562 (98.6)	<0.001
Yes	9 164	8 430 (0.4)	734 (1.4)	
NND[a]				
No	1 99 9270	1 947 791 (99.8)	51 479 (98.4)	<0.001
Yes	5 317	4 500 (0.2)	817 (1.6)	
Gestational age (weeks)				
24-28	10 280	8 233 (0.4)	2 047 (3.9)	<0.001
29-32	22 211	17 550 (0.9)	4 661 (8.9)	
33-34	29 080	22 851 (1.2)	6 229 (11.9)	
35-36	78 915	65 728 (3.4)	13 187 (25.2)	
37-38	327 284	326 227 (16.7)	21 057 (40.3)	
39->40	1 516 817	1 511 702 (77.4)	5 115 (9.8)	
Missing	0			
Maternal age (years)				
15-20	229 683	226 654 (11.6)	3 029 (5.8)	<0.001
21-24	378 933	372 028 (19.1)	6 905 (13.2)	
25-29	743 004	727 925 (37.3)	15 079 (28.8)	
30-34	455 549	438 880 (22.5)	16 669 (31.9)	
35-39	172 580	163 638 (8.4)	8 942 (17.1)	
>40	24 838	23 166 (1.2)	1 672 (3.2)	
Missing	0			
Year of birth				
1980-1985	310 806	303 740 (15.6)	7 066 (13.5)	<0.001
1986-1990	317 353	310 417 (15.9)	6 936 (13.3)	
1991-1995	312 276	304 636 (15.6)	7 640 (14.6)	
1996-2000	276 922	269 451 (13.8)	7 471 (14.3)	
2001-2005	251 658	244 293 (12.5)	7 365 (14.1)	
2006-2010	278 059	269 421 (13.8)	8 638 (16.5)	
2011-2015	257 513	250 333 (12.8)	7 180 (13.7)	
Missing				
Parity				
Primigravida	895 000	871 589 (44.6)	23 411 (44.8)	0.58
Parous	1 109 587	1 080 702 (55.4)	28 885 (55.2)	
Missing	0			
Fetal Sex				
Male	1 027 819	1 001 632 (51.3)	26 187 (50.1)	<0.001
Female	979 768	950 659 (48.7)	26 109 (49.9)	
Missing	0			
Social deprivation category				
1	340 182	330 031 (17.2)	10 151 (19.8)	
2	351 277	341 857 (17.8)	9 420 (18.3)	<0.001
3	386 822	376 807 (19.7)	10 015 (19.5)	
4	409 798	399 609 (20.8)	10 189 (19.8)	
5	481 326	469 703 (24.5)	11 623 (22.6)	
Missing	35 182			
Maternal smoking at booking				
Smoker	33 0952	323 528 (16.6)	7 424 (14.2)	<0.001
Non-smoker	629 686	603 000 (30.9)	26 686 (51.03)	
Missing	1 043 949			
Birth-weight centiles				
<10	204 160	198 840 (10.2)	5 320 (10.2)	0.961
11-97	1 739 488	1 694 328 (86.9)	45 160 (86.9)	
>97	58 461	56 947 (2.9)	1 514 (2.9)	
Missing	2 478			

[a]NND = Neonatal death, [b]N = Number

8.4.1 Stillbirth in Twins Compared to Singletons According to Gestation at Birth

The primary outcome of stillbirth by week of gestation in twins compared to singletons is shown in Figure 8-2. Compared to singleton infants, twin infants born at every week of gestation from 24 weeks had increased odds of stillbirth and this difference was most marked at 38 weeks gestation (adj. OR 14.97, 95% CI 10.71-20.92). Table 8-2 summarises the results by gestational age category. The highest odds of stillbirth in twins compared with singletons was observed for early term births (37-38 weeks, adj. OR 18.06, 95% CI 14.26-22.88).

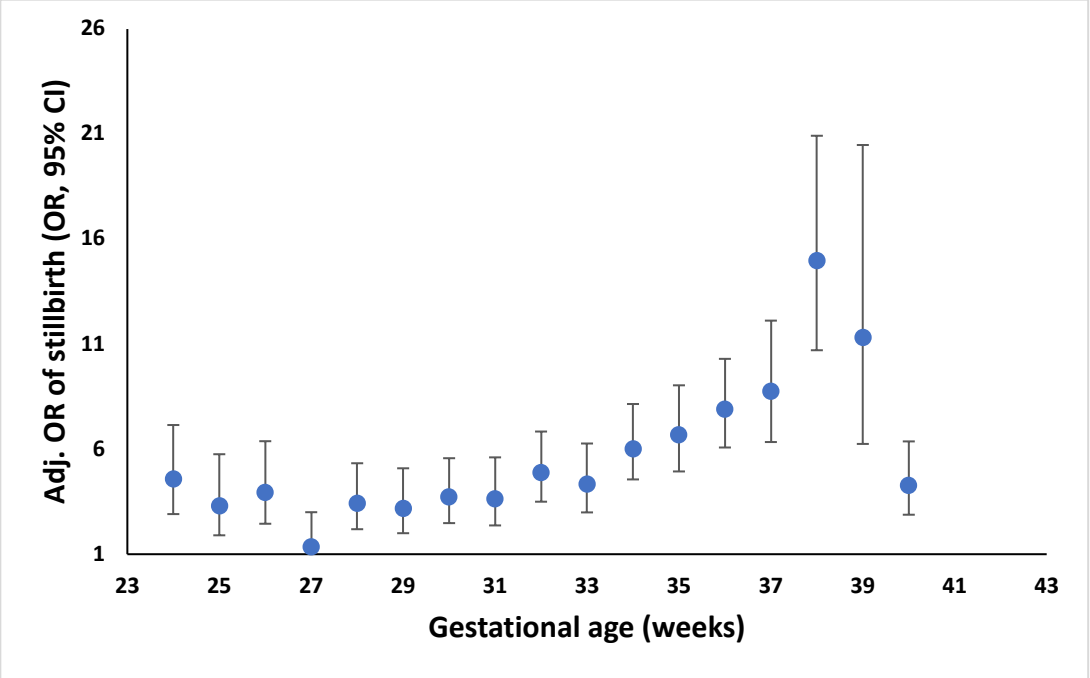


Figure 8-2: Adjusted odds of stillbirth in twins compared to singletons in each gestational week from 24 weeks (adjusted for maternal age, parity, fetal sex, year of delivery, social deprivation category and birth weight centiles)

Table 8-2: Odds of stillbirth at each gestational age category in twins compared to singletons

	Twins N[a] with stillbirth/ongoing pregnancies	Singletons N with stillbirth/ongoing pregnancies	Unadjusted OR[b] (95% CI[c])	P value	Adjusted OR* (95% CI)	P value
Extreme preterm <28 weeks	167/50249	1728/1944058	3.74 (3.08-4.54)	<0.001	3.33 (2.65-4.19)	<0.001
Very preterm 29-32 weeks	168/45588	1599/1926508	4.44 (3.69-5.35)	<0.001	4.03 (3.30-4.93)	<0.001
Moderate preterm 33-34 weeks	114/39359	846/1903657	6.52 (5.29-8.03)	<0.001	5.40 (4.27-6.82)	<0.001
Late preterm 35-36 weeks	140/26172	1052/1837929	9.34(7.76-11.25)	<0.001	8.08 (6.62-9.85)	<0.001
Early term 37-39 weeks	109/5115	1271/1551702	25.35 (20.70-31.03)	<0.001	18.06 (14.26-22.88)	<0.001
Full term 39->42 weeks	36/5079	1934/1509768	5.53 (3.91-7.83)	<0.001	4.28 (2.89-6.37)	<0.001

[a]N = Number, [b]OR = Odds ratio, [c]CI = 95% Confidence intervals *adjusted for maternal age, parity, baby sex, year of delivery, deprivation category and birth weight centiles

Table 8-3: Odds of NND at each gestational age category in twins compared to singletons

	Twins N[a] with NND/live births	Singletons N with NND/live births	Unadjusted OR[b] (95% CI[c])	P value	Adjusted OR* (95%CI)	P value
Extreme preterm <28 weeks	547/1333	1517/4934	1.29 (1.13-1.47)	<0.001	1.39 (1.20-1.61)	<0.001
Very preterm 29-32 weeks	134/4359	661/15290	0.71 (0.58-0.88)	<0.001	0.78 (0.63-0.97)	0.024
Moderate preterm 33-34 weeks	35/6080	300/21705	0.41 (0.29-0.60)	<0.001	0.48 (0.33-0.69)	<0.001
Late preterm 35-36 weeks	48/12999	362/64314	0.66 (0.48-0.91)	<0.001	0.71 (0.52-0.99)	0.046
Early term 37-39 weeks	40/20908	468/324488	1.33 (0.95-1.85)	<0.001	1.25 (0.89-1.77)	0.198
Full term 39->42 weeks	13/5066	1138/1508630	3.40 (1.97-5.88)	<0.001	3.00 (1.73-5.18)	<0.001

[a]N = Number, [b]OR = Odds ratio, [c]CI = 95% Confidence intervals *adjusted for maternal age, parity, baby sex, year of delivery, deprivation category and birth weight centiles

8.4.2 Neonatal Death in Twins Compared to Singletons According to Gestation at Birth

Compared to singletons, NND rates were higher in twins in the extreme preterm period (<28 weeks adj. OR 1.39, 95% CI 1.20-1.61, Table 8-3) and for births beyond 39 weeks gestation (adj. OR 3.00, 95% CI 1.73-5.18). However, there was no significant difference for birth at 37-39 weeks (adj. OR 1.25, 95% CI 0.89-1.77) and NNDs were less likely to occur in twins compared to singletons for all preterm gestational ages other than extreme preterm (adj. OR 0.78, 95% CI 0.63-0.97) at 29-32 weeks; adj. OR 0.48, 95% CI 0.33-0.69 at 33-34 weeks and adj. OR 0.71, 95% CI 0.52-0.99, Table 8-3).

8.4.3 Subgroup Analysis of Non-medically Indicated Births

In the planned subgroup analysis, excluding medically-indicated deliveries, the results were more marked for stillbirth. There were increased odds of stillbirth in twins with non-medically indicated birth compared to the results for all twins compared to singletons (most marked difference at 37-39 weeks adj. OR 19.26, 95% CI 14.90-24.90 compared to an adj. OR of 18.06 (14.26-22.88), <0.001 in all twins compared to singletons, Table 8-4). For NND, the results in the subgroup excluding the medically indicated births were very similar to the results for all twins compared to singletons (Table 8-4).

Table 8-4: Subgroup analysis of non-medically indicated deliveries and their relationship with stillbirth and NND (n=1,885,923 for non-medically indicated deliveries)

Stillbirth		
Week of gestation	All Twins compared to singletons Adj. OR (95% CI), p value	Non-medically indicated twin deliveries compared to singletons, Adj. OR (95% CI), p value
Extreme preterm <28 weeks	3.33 (2.65-4.19), <0.001	4.23 (3.37-5.34), <0.001
Very preterm 29-32 weeks	4.03 (3.30-4.93), <0.001	5.01 (4.01-6.20), <0.001
Moderate preterm 33-34 weeks	5.40 (4.27-6.82), <0.001	6.35 (4.96-8.16), <0.001
Late preterm 35-36 weeks	8.08 (6.62-9.85), <0.001	9.14 (7.37-11.33), <0.001
Early term 37-39 weeks	18.06 (14.26-22.88), <0.001	19.26 (14.90-24.90), <0.001
Full term 39->42 weeks	4.28 (2.89-6.37), <0.001	4.76 (3.17-7.14), <0.001
NND		
Extreme preterm <28 weeks	1.39 (1.20-1.61), <0.001	1.39 (1.20-1.61), <0.001
Very preterm 29-32 weeks	0.78 (0.63-0.97), 0.024	0.75 (0.60-0.93), 0.010
Moderate preterm 33-34 weeks	0.48 (0.33-0.69), <0.001	0.43 (0.29-0.63), <0.001
Late preterm 35-36 weeks	0.71 (0.52-0.99), 0.046	0.64 (0.45-0.92), 0.015
Early term 37-39 weeks	1.25 (0.89-1.77), 0.198	1.44 (1.01-2.05), 0.043
Full term 39->42 weeks	3.00 (1.73-5.18), <0.001	3.29 (1.90-5.69), <0.001

8.4.4 Subgroup Analysis of Sex Discordant (Certain to be Dichorionic) Twins

Results of this planned subgroup analysis of sex discordant twins were similar to the results for all twins; twin infants had increased odds of stillbirth compared to singleton pregnancies for each week of gestation and this difference was most marked at 37-39 weeks gestation (adj. OR 18.87, 95% CI 12.82-27.78, Appendix Table 8-2). Sex discordant twins, similar to all twins, had higher odds of NND compared to singletons in the extreme preterm period (adj. OR 1.66, 95% CI 1.30-2.13 <28 weeks gestation) and in deliveries greater than 39 weeks (adj. OR 5.15, 95% CI 2.45-10.83, Appendix Table 8-3) but lower for all preterm gestational ages other than extreme preterm.

8.4.5 Sensitivity Analysis

The results were similar to the main results presented in Tables 8-2 and 8-3 when we ran the logistic regression models including maternal smoking as a covariate in the sub-group of women in which this information was available (Appendix Table 8-4).

8.5 Discussion

Contrary to the results of previous studies reporting a lower risk of stillbirth in preterm born twins compared to singletons (Vasak et al. 2017, Kahn et al. 2003, Minakami and Sato 1996), we found that twins born at all gestational ages had a greater odds of stillbirth compared to singletons after adjustment for known potential confounders including medical indication for delivery. The pattern was different in the case of NND where we found higher odds of NND in twins (compared with singletons) in the early preterm period (less than 28 weeks gestation) and in deliveries after 39 weeks but lower odds of NND in twins born in the gestational age categories between 29 and 37 weeks gestation compared to singletons.

The lower NND rates in twins compared to singletons at preterm gestations could relate to the different aetiologies of the preterm delivery affecting twins and singletons. Preterm birth results from a number of aetiologies, and contributing causes are likely to be proportionately different in twins and singletons. The different pathological processes leading to preterm birth are also likely to predominate at different gestations, and thus these may influence relative survival between twins and singletons. For example, birth weight is an important determinant of neonatal outcome and twins are eight times more likely to have a low birthweight compared to singletons (Min et al. 2000). In the extreme preterm period this may explain the higher NND rates in twins compared the singletons as although all babies born at this gestation are small, the twins are more likely to be low birthweight for gestation than

singletons born in the same gestational age category. Another example is uterine stretch as the main cause of preterm birth in twins compared to pathological processes, namely infection, in singletons which may lead to a lower NND rate in twins in the preterm gestational age categories. The lower NND rates in twins compared to singletons for delivery at 28-37 weeks gestation was unexpected. One other potential explanation for this is different neonatal care between twins and singletons. This would only be a difference in the preterm period however as the risks are higher in twins after 39 weeks and therefore a more plausible explanation is the one given above that twins born preterm are 'born in a better condition' than singletons born preterm. It is important to note that likely due to the policy of planned delivery of twins from 37-38 weeks, our sample of twins in the >39 week gestational age bracket was small and the results for this gestational age category should be interpreted with caution.

Our findings from this study show that the most marked increase in stillbirth between twins and singletons occur at the 38 week gestation mark and therefore fits with the need for more intense monitoring of twin pregnancies compared to singleton pregnancies. Twins and singletons are managed differently antenatally with twin pregnancies regarded as 'high risk', frequently attending obstetric antenatal clinics and undergoing regular growth ultrasound scans (Kilby 2017). In low-risk singleton pregnancies in the UK, pregnant women have largely midwifery-led care with no further ultrasound scans other than the routine booking and anomaly scans (NICE 2008). It is not entirely clear why twins have an increased risk of stillbirth compared to singletons and why this occurs at an earlier gestation but it is likely due to a form of accelerated maturation of the placenta (Fuchs and Senat 2016). Our findings go against the theory presented in previous studies (Vasak et al. 2017) that twin pregnancies are monitored more closely resulting in lower preterm perinatal death as we found an increased risk of stillbirth at each week of gestation in twins compared to singletons despite the increased antenatal monitoring of twin pregnancies. These findings justify the increased antenatal monitoring of twin pregnancies compared to singletons but suggests that this increased monitoring may be working in lowering the NND rates at some preterm gestations but fails to completely mitigate the excess risk.

A strength of this study is the large, unselected twin and singleton population sample. Using routinely collected population maternity data reduced the risk of selection, reporting and recall bias. The study also had a number of methodological advantages over previous studies. Firstly, we were able to account for the clustering of the twin infants using GEE analysis, this is important as if clustering is not accounted for it can lead to an overestimate of the effect. We also analysed the primary outcomes of stillbirth and NND separately using a denominator of ongoing pregnancies at risk for stillbirth and livebirths at risk for NND hence we could be sure the correct denominators were being used in the analysis. Overall calculating the risk of stillbirth and NND in the conventional way using a denominator of all births, twins had higher odds of stillbirth (OR 3.28, 95% CI 3.04-3.54) and NND (OR 6.87, 95% CI 6.37-7.40) compared to singletons (Appendix Table 1). Analysing stillbirth and NND separately as opposed to studying perinatal mortality rate (PMR) explains the different results reported in previous studies (Vasak et al. 2017) and we believe this is a more accurate way of performing and reporting the analyses with the correct denominators.

Due to the nature of routinely collected data there are some limitations of the study. If missing data are not missing at random, this can lead to selection bias if they relate to the inclusion or exclusion criteria or residual confounding if they related to covariates not included in the multivariate models. Missing data can also reduce study power. For example, missing data on maternal smoking was a potential source of residual confounding; however, a sensitivity analysis including maternal smoking in a sub-group of women produced similar results to the overall. Another limitation of the study was being unable to adjust or stratify by chorionicity. Chorionicity (number of placentae in a twin pregnancy) is an important risk factor for perinatal death in twins with monochorionic twins (shared placenta with either two separate chorions [external fetal membranes, monochorionic diamniotic] or shared chorion [monochorionic monoamniotic]) having a two-fold greater perinatal mortality compared to dichorionic twins (two placentae and two separate chorions)(Kilby 2017). We therefore performed a planned subgroup analysis of sex discordant twins

to represent dichorionic twins and the results were very similar to the results for all twins with an increased risk of stillbirth in twins at each week of gestation and a higher NND in extreme preterm and in twins born greater than 39 weeks compared to singletons.

In conclusion, this study shows that, overall twins have an increased risk of stillbirth and neonatal death compared to singleton pregnancies. In contrast to previous studies, but in line with what would be expected clinically, the odds of stillbirth were higher in twins at every week of gestation compared to singletons and this was most marked at 38 weeks. NND odds were higher in twins in the extreme preterm period and in births greater than 39 weeks gestation compared to singletons but was lower in twins from 28-37 weeks and this may be due to the different contributing causes of the preterm birth in twins and singletons

8.6 Appendices

Appendix Table 8-1: Overall risk of perinatal death, stillbirths and NNDs in twins compared to singletons

Outcome	N[a] twins	N singletons	Crude OR[b] (95% CI[c])	P value
Perinatal Death	1551/52296	12930/1952291	4.58 (4.35-4.84)	<0.001
Stillbirth	734/52296	8430/1952291	3.28 (3.04-3.54)	<0.001
Neonatal Death	817/50745	4500/1939361	6.87 (6.37-7.40)	<0.001

[a]N = number, [b]OR = Odds ratio, [c]CI = 95% Confidence intervals

Appendix Table 8-2: Odds of Stillbirth at each gestational age category in sex discordant twins compared to singletons

	Twins N[a] with stillbirth/ongoing pregnancies	Singletons N with stillbirth/ongoing pregnancies	Unadjusted OR[b] (95% CI[c])	Adjusted OR* (95% CI)	P value
Extreme preterm <28 weeks	15/16614	1728/1944058	1.02 (0.59-1.75)	0.82 (0.41-1.61)	0.558
Very preterm 29-32 weeks	16/15271	1599/1926508	1.26 (0.75-2.13)	1.25 (0.73-2.15)	0.422
Moderate preterm 33-34 weeks	26/13373	846/1903657	4.37 (2.92-6.56)	4.23 (2.79-6.41)	<0.001
Late preterm 35-36 weeks	40/9162	1052/1837929	7.63 (5.55-10.47)	6.89 (4.94-9.61)	<0.001
Early term 37-39 weeks	35/1659	1271/1511702	25.01 (17.81-35.35)	18.87 (12.82-27.78)	<0.001
Full term 39->42 weeks	16/1643	1934/1509768	7.60 (4.65-12.43)	4.79 (2.58-8.90)	<0.001

[a]N = Number, [b]OR = Odds ratio, [c]CI = 95% Confidence intervals *adjusted for maternal age, parity, baby sex, year of delivery, deprivation category and birth weight centiles

Appendix Table 8-3: odds of NND at each gestational age category in sex discordant twins compared to singletons

	Twins N[a] with NND/live births	Singletons N with NND/live births	Unadjusted OR[b] (95% CI[c])	Adjusted OR* (95% CI)	P value
Extreme preterm <28 weeks	140/346	1571/4934	1.27 (1.00-1.62)	1.66 (1.30-2.13)	<0.001
Very preterm 29-32 weeks	32/1295	6661/15290	0.57 (0.40-0.82)	0.67 (0.47-0.96)	0.030
Moderate preterm 33-34 weeks	10/1862	300/21705	0.39 (0.21-0.73)	0.45 (0.24-0.85)	0.014
Late preterm 35-36 weeks	9/4162	362/64314	0.38 (0.20-0.74)	0.48 (0.24-0.92)	0.028
Early term 37-39 weeks	15/7453	468/324488	1.40 (0.83-2.33)	1.54 (0.90-2.62)	0.112
Full term 39->42 weeks	7/1652	1138/1510564	5.63 (2.68-11.82)	5.15 (2.45-10.83)	<0.001

[a]N = Number, [b]OR = Odds ratio, [c]CI = 95% Confidence intervals *adjusted for maternal age, parity, baby sex, year of delivery, deprivation category and birth weight centiles

Appendix Table 8-4: Sensitivity analysis: Results displayed with and without including maternal smoking as a covariate

Stillbirth		
Week of gestation	All Twins compared to singletons Adjusted OR (95% CI), p value	With maternal smoking; twins compared to singletons, Adjusted OR (95% CI), p value
Extreme preterm <28 weeks	3.33 (2.65-4.19), <0.001	2.90 (2.11-3.98), <0.001
Very preterm 29-32 weeks	4.03 (3.30-4.93), <0.001	3.62 (2.66-4.94), <0.001
Moderate preterm 33-34 weeks	5.40 (4.27-6.82), <0.001	5.34 (3.79-7.53), <0.001
Late preterm 35-36 weeks	8.08 (6.62-9.85), <0.001	8.07 (6.62-9.85), <0.001
Early term 37-39 weeks	18.06 (14.26-22.88), <0.001	23.93 (15.61-36.66), <0.001
Full term 39->42 weeks	4.28 (2.89-6.37), <0.001	2.90 (1.21-6.96), 0.017
NND		
Extreme preterm <28 weeks	1.39 (1.20-1.61), <0.001	1.49 (1.16-1.91), 0.002
Very preterm 29-32 weeks	0.78 (0.63-0.97), 0.024	0.84 (0.54-1.31), 0.444
Moderate preterm 33-34 weeks	0.48 (0.33-0.69), <0.001	0.37 (0.18-0.82), 0.014
Late preterm 35-36 weeks	0.71 (0.52-0.99), 0.046	1.32 (0.78-2.21) 0.299
Early term 37-39 weeks	1.25 (0.89-1.77), 0.198	1.41 (0.82-2.40), 0.212
Full term 39->42 weeks	3.00 (1.73-5.18), <0.001	3.29 (1.90-5.69), <0.001

8.7 Chapter Conclusions

The work presented in **Chapter 8** demonstrated that overall twins have increased odds of stillbirth compared to singletons and contrary to previous studies, this rate was increased at every week of gestation from 24-42 weeks in twins compared to singletons. Neonatal death was also higher overall in twins compared to singletons but was lower in 29-27 weeks of gestation. Further research is necessary to determine why this difference occurs but it is likely due to the different aetiologies of preterm birth between twins and singletons.

Chapter 8 explored the use of routinely collected population maternity data to compare outcomes in twins compared to singletons. **Chapter 7** looked at the long-term childhood outcomes of twins according to gestation at birth using routinely collected education data linked to maternity data. In **Chapter 9**, the long-term outcomes of singletons according to gestation at birth were studied in the form of a systematic review of previously conducted studies.

Chapter 9

Long term cognitive outcomes of early term (37-38 weeks) and late preterm (34-36 weeks) births: A systematic review

The following materials have been published in Wellcome Open Research in 2018 (Murray et al., 2018) under the same title by Dr Sarah R Murray (SM), Dr Susan Shenkin (SS), Dr Kirsten McIntosh (KM), Dr Jane Lim (JL), Benjamin Grove (BG), Professor Jill Pell (JP), Professor Jane Norman (JN) and Dr Sarah Stock (SJS). SS and SJS conceived the study and instigated the collaboration. SS and JL designed the study protocol and registered the review on PROSPERO (The International Prospective Register of Systematic Reviews). SM, JL, KM and BG conducted the searches, selection of studies, assessment of bias and data extraction. SM wrote the first draft of the manuscript under the guidance of SS and SJS. All authors provided critical insight for the final draft of the manuscript and approved the final submitted article.

In summary, this work demonstrated that singletons born at 39-41 weeks gestation had higher cognitive scores at school than those born at 37-38 weeks and those born after 42 weeks gestation. The review found no differences in cognitive outcomes in children born late preterm (34-36 weeks) compared to children who were born at term (37-43 weeks).

This work concluded that although the differences in IQ scores between children born at early term (37-38 weeks) compared to full term (39-41 weeks) were small (approximately 3 IQ points) this may not be important at an individual level but at a population level, when the gestational age at delivery has been reducing, this may have marked consequences. Due to the heterogeneous nature of the studies included in terms of cognitive scores used and outcomes reported, a meta-analysis was not possible and instead a narrative review was undertaken.

9.1 Abstract

Background: There is a paucity of evidence regarding long-term outcomes of late preterm (34-36 weeks) and early term (37-38 weeks) delivery. The Objective of this systematic review was to assess long-term cognitive outcomes of children born at these gestations.

Methods: Four electronic databases (Medline, Embase, clinicaltrials.gov and PsycINFO) were searched. Last search was 5th August 2016. Studies were included if they reported gestational age, IQ measure and the ages assessed. The protocol was registered with the International prospective register of systematic reviews (PROSPERO Record CRD42015015472, http://www.crd.york.ac.uk/PROSPERO/). Two independent reviewers assessed the studies. Data were abstracted and critical appraisal performed of eligible papers.

Results: Of 11,905 potential articles, seven studies reporting on 41,344 children were included. For early term births, four studies (n = 35,711) consistently showed an increase in cognitive scores for infants born at full term (39-41 weeks) compared to those born at early term (37-38 weeks) with increases for each week of term (difference between 37 and 40 weeks of around 3 IQ points), despite differences in age of testing and method of IQ/cognitive testing. Four studies (n = 5644) reporting childhood cognitive outcomes of late preterm births (34 – 36 weeks) also differed in study design (cohort and case control); age of testing; and method of IQ testing and found no differences in outcomes between late preterm and term births although risk of bias was high in included studies.

Conclusion: Children born at 39-41 weeks have higher cognitive outcome scores compared to those born at early term (37-38 weeks). This should be considered when discussing timing of delivery. For children born late preterm the data is scarce and when compared to full term (37-42 weeks) did not show any difference in IQ scores.

9.2 Introduction

Globally preterm birth rates are rising with 10% of neonates born less than 37 weeks gestation (Blencowe et al. 2012). Late preterm births (34-36 weeks) account for three quarters of all preterm births (Blencowe et al. 2013). Early term births (37-38 weeks gestation) have also increased and contribute substantially to an overall decrease in gestational age at delivery. In the US the average gestational age at delivery has decreased from 40 weeks in 1994 to 39 weeks of gestation in 2004 (Gyamfi-Bannerman 2011).

Early term delivery is associated with increased short term adverse physical morbidity including respiratory distress syndrome, transient tachypnoea of the neonate and ventilator use as well as an increased risk of infant mortality at 37 weeks compared to full-term delivery(Reddy et al. 2009, Vohr 2013, Martínez-Nadal et al. 2014). It is for this reason that both the Royal College of Obstetricians and Gynaecologists UK (RCOG, 2010) and the American college of Obstetricians and Gynecologists (ACOG, 2013)endorse the policy of elective birth after 39 weeks in order to reduce the risk of adverse outcome in infants born before full term (39-40 weeks gestation). There is a paucity of evidence regarding the long-term morbidity of this group, in particular the impact on cognitive function. Advanced gestational age is associated with a lower risk of having special educational need at school (MacKay et al. 2010). Davis *et al.*(Davis et al. 2011) has also shown that even amongst the weeks of term advanced gestational age is associated with better neurodevelopment as demonstrated by magnetic resonance imaging (MRI). As obstetric efforts worldwide continue to attempt to reduce stillbirth amongst term deliveries, induction of labour at an earlier gestational age is becoming more common, despite the guidance above, and therefore it is imperative to consider the long-term outcomes of deliveries before term to guide clinicians and parents on optimum timing of delivery.

The association between preterm birth and long-term neurological morbidity is better established with the risk increasing with decreasing gestational age with extremely

preterm babies (≤ 26 weeks) having the worst neurological outcomes (Moore et al. 2012, Costeloe et al. 2012). The aetiology of this is hypothesized to be due to the disruption of the pathways of dendritic arborisation, synaptogenesis and the thickening of the developing cortex (Huttenlocher and Dabholkar 1997). However, there is less evidence regarding long-term cognitive outcomes of late preterm/early term infants and given they account for the largest proportion of singleton preterm births more research is necessary. A systematic review of 29,375,675 late preterm infants (34-36 weeks)(Teune et al. 2011), found increased risks of cerebral palsy (RR 3.1, 95% CI 2.3-4.2) and lower likelihood of finishing school in the late preterm born infants (RR 0.96, 95% CI 0.95-0.97) but we could find no prior reviews on cognitive outcomes for early term births.

The aims of this systematic review are to describe the objectively measured cognitive outcomes in childhood up to the age of 18 years i) within each gestational week of term (37-42 weeks) and ii) of late-preterm (34-36 weeks) compared to term (37-42 weeks) deliveries. The results are necessary for informed decision-making regarding timing of delivery.

9.2 Methods

This systematic review of the literature was conducted according to the STROBE guidelines (Von Elm et al. 2007) and reported according to the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al. 2009). The study protocol was registered with the University of York Centre for Reviews and Dissemination International prospective register of systemic reviews (PROSPERO Record CRD42015015472, http://www.crd.york.ac.uk/PROSPERO/). MEDLINE (1946-2016), EMBASE (1947-2016) and PsycINFO (1945-2016) were searched using a search strategy developed and tested in collaboration with a librarian experienced in literature searching (Appendix 1). The searches were supplemented with a manual search through the reference lists of selected primary articles. A forward citation search was

performed on all included studies. The first search date was 12th January 2015 and the last search was 5th of August 2016. A subsequent search on Clinicaltrials.gov was performed on 2nd June 2017.

9.2.1 Study Selection

One reviewer (JL) screened all titles and abstracts and a second reviewer (BG) independently screened through a 10% sample of the 10,882 articles by reading the title and abstracts of the first 100 articles of every 1000. The search was updated in August 2016 which yielded an additional 1023 titles and abstracts screened independently by two reviewers (SM, KM). After a consensus was reached the full texts were retrieved and critically appraised by both reviewers independently (SM, KM). We contacted the individual authors of the included studies to obtain the data necessary to complete the results table. Reasons for exclusion were recorded.

Late-preterm birth was defined as a live birth from 34 to 36+6 completed weeks of gestation (Chang et al. 2013). The primary outcome was the results of standardised general intelligence quotient (IQ) tests before age 18 rather than specific domains of cognition. General cognitive ability of a physically and neurologically normal, healthy population of individuals was the key outcome measure recorded. Term birth was defined as a live birth from 37 to 42 completed weeks of gestation.

Studies were included if they reported the range of the participants gestational age, assessment of IQ using a validated score; and the age when IQs was assessed. There were no restrictions by study design, language or method of gestational age assessment. Preterm participants were included as long as there was a clear subgroup of gestational age of 34-36 weeks. Excluded studies included those with: unclear method of cognitive testing; if only selected domains of cognition (e.g. verbal intelligence) were tested; if educational outcomes rather than IQ reported; studies involving high-risk or atypical groups as controls (e.g. multiple births, intra-uterine

growth restriction, those with bronchopulmonary dysplasia or brain haemorrhage). Full details of inclusion and exclusion criteria are presented in Appendix 2.

The quality of studies was assessed based on the representativeness of the general population, the method of measurement of gestational age and the recording of intelligence testing using the Risk of Bias Assessment Tool for nonrandomized studies (RoBANS) tool (Kim et al. 2013)(Appendix 3). Two independent reviewers extracted data from each paper on study location, design, population, IQ score used and the main results.

The studies differed widely in outcome measures of cognition used and due to the large heterogeneity between the study designs and methods a meta-analysis was not possible.

9.3 Results

The database and additional record search identified 11,905 articles after removal of duplicates. After exclusions (See flow diagram, Figure 9.1) six studies and one conference abstract (which reported on both late preterm and outcomes within term and is therefore included in both groups) reporting on 41,344 children/adolescents were included in the review; four studies comparing the outcomes within term (37-42 weeks) and four studies comparing the outcomes of late preterm delivery (34-36 weeks) with term delivery.

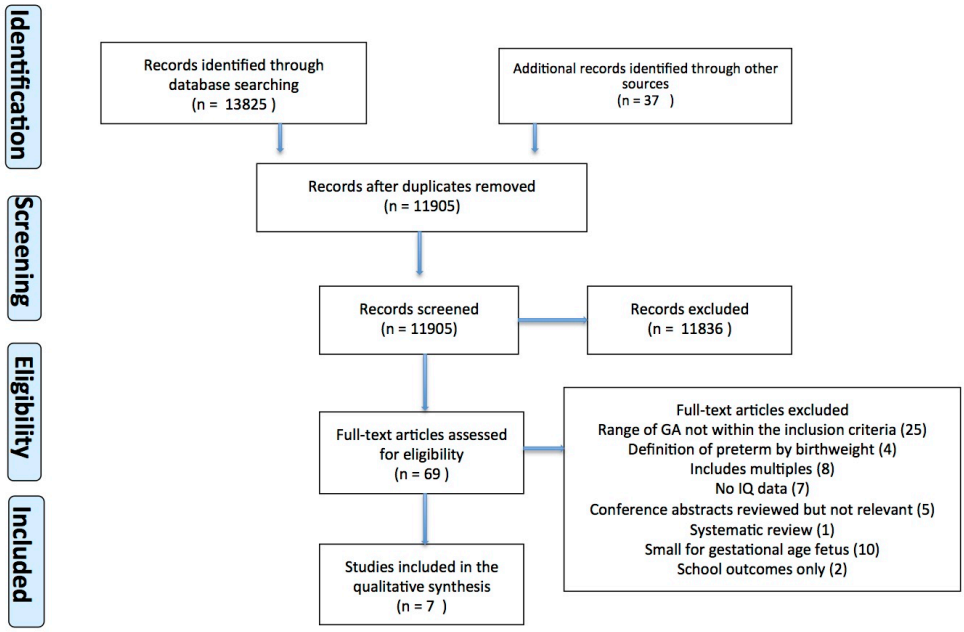


Figure 9-1: Flowchart showing the study selection process (adapted from the PRISMA flow diagram)

The studies comparing the outcomes within term differed in a number of ways including: age of testing (1 year, 4 years and 6 years); method of IQ testing (Bayley Scales of Infant Development (BSID), the Stanford-Binet general IQ test and the Wechsler Abbreviated Scale of Intelligence (WASI); details in Appendix 4); and the categories of gestational age investigated (37-41 weeks in one study, 37-42 weeks in two studies and 37-43 weeks in one study).

The studies comparing the outcomes of late preterm delivery (34-36 weeks) also differed in a number of ways including: study design (three prospective cohort studies and one case control study); age of testing (2 years and 13-14 years); and the method of IQ testing (Bayley Scales of Infant Development (BSID) and the Wechsler Intelligence Scale for Children (WISC); details in Appendix 4).

Table 9-1 provides details on the characteristics of the included studies. Three studies (Espel et al. 2014, Rose et al. 2013, Yang, Platt and Kramer 2010) and one study which was only available as a conference abstract (on contacting the author no further information was available as it is not yet published)(Gyamfi-Bannerman, Son and Ananth 2014) reporting on term deliveries (35,711 children/adolescents) and three studies (Nepomnyaschy et al. 2012, Romeo et al. 2016, Narberhaus et al. 2007)(as well as the conference abstract) on late preterm deliveries (5,644 children/adolescents).

Table 9-1 Characteristics of the included studies, ordered by gestational age categories and age of participant testing.

Author, Year (country)	Study design, total n	GA (n)	Control (n)	Age assessed (yrs)	Exclusions	Exposure measurement	Cognitive outcomes*	Adjustment
Comparison within term (37-42 weeks)								
Espel et al. 2014[18] (USA)	Longitudinal prospective study N = 232	37-41 (232)	-	1	Preterm birth (<37 weeks), Post term birth (>41 weeks)	GA measured from LMP And Ultrasound	BSID II assessed by trained clinicians, no blinding	Maternal ethnicity; Marital status; Fetal Sex; Obstetric complications; Maternal education; Parity; Socio-economic status; Birth weight percentile
Rose et al. 2013[19] (Chile)	Prospective cohort from an RCT N = 1,562	37-42 (1562)	-	1	Infants born at >42 weeks gestation, birth weight <3kg	GA based on LMP	BSID II performed by trained psychologists, blinding not reported	Fetal sex Socio-economic status Birthweight
Gyamfi Bannerman et al. 2014[21] (USA)	Prospective cohort study N = 20,093	37-42 (18,242)		4	Parents with mental retardation or alcohol or drug use	Not reported	Full-scale IQ scores from the Stanford-Binet IQ	Maternal age; Maternal education; Socio-economic status; Smoking
Yang et al. 2009[20] (Belarus)	Prospective cohort from an RCT N = 13,824	37 (469) 38 (2100) 42 (171) 43 (10)	39-41 (11, 074)	6.5	Birth weight <2500g	GA from hospital records, 94% confirmed by ultrasound	Full-scale IQ from WASI scores performed by trained pediatricians	Maternal age; Marital status; Maternal education; Parity Smoking; Maternal height; Parental occupation
Late preterm vs term (34 – 36+6)								
Nepomnyaschy et al. 2011[22] (USA)	Prospective cohort study N = 5450	34-36 (400)	37-41 (5050)	2	Mothers <15 yrs, congenital anomalies, multiple births	GA measured by maternal estimated LMP	General, BSID II (converted to a percentile) mental ability, Language by trained interviewer blinded to GA	Maternal ethnicity; Maternal age; Marital status; Fetal sex Obstetric complications; Maternal Education; Parity Socio-economic status; Smoking
Romeo et al. 2015[23] (Italy)	Prospective cohort study N -= 119	34-36 (71)	37-41 (48)	2	Abnormal cranial Ultrasound, congenital malformations, SGA	Not reported	MDI of BSID II at 2 yrs Blinding and interviewer training not reported	Fetal sex
Narberhaus et al. 2007[24] (Spain)	Case Control study N = 64	34-36 (11)	37-41 (53)	13-14	Mental or physical disability, perinatal complications/hypoxic risk	Not reported	General WISC full scale assessed by the same psychologist	Matched by: Maternal ethnicity; Maternal age; Fetal sex Maternal education; Socio-economic status

*Cognitive Outcomes (details in appendix 3): BSID-II = Bayley Scales of Infant Development (2nd Edition); MDI = Mental Developmental Index, PDI = Psychomotor Developmental Index, WISC = Wechsler Intelligence Scale for Children, WASI = Wechsler Abbreviated Scale of Intelligence

Table 9-2 shows the cognitive outcomes of each week of gestation among term births (range 37-43 weeks). In general, although the studies differed in age at assessment and the IQ test used, all four studies (Espel et al. 2014, Rose et al. 2013, Yang et al. 2010, Gyamfi-Bannerman et al. 2014) showed an increase in cognitive outcome scores for Infants born at full term (39-41 weeks) compared to those children born at early term (37-38 weeks) with statistically significant increases for each week of term. Three of the studies were at classed as moderate risk of bias (Espel et al. 2014, Rose et al. 2013, Yang et al. 2010) and one was at high risk because it did not have any clear information on how gestational age was measured (Gyamfi-Bannerman et al. 2014). Yang *et al.*(Yang et al. 2010) was the only study to measure outcomes up to post-term (43 weeks gestation) and found an inverse U-shaped relationship of IQ score and gestational age. In this large study (n = 13,824), with a moderate risk of bias, full-term (39-41 weeks) was used as a reference group with mean differences in IQ scores reported at early term (37-38 weeks) weeks which were lower than full term and also post term (42-43 weeks) which had a higher mean difference from full term than early term (for full risk of bias see appendix 3). The effect size cannot be summarised due to differences between the studies but the IQ difference between 37 and 40 weeks was approximately 3 IQ points. This may not be clinically significant at an individual level, but would have an impact at a population level.

Table 9-2 Results of individual studies comparing cognitive outcomes of children born within term (37-42 weeks gestation), ordered by participant age at cognitive testing.

Study (n)	Cognitive test and overall scores	Age at testing (years)	Main findings Scores	Effect size (95%CI) /significance level		Risk of bias[a]
Espel *et al.* 2014[18] (n = 232)	BSID-II (Bayley scores of infant development) Range 50-150 MDI and PDI[e]	1	GA[b] / MDI mean ± SD[c] / PDI mean±SD[d] 37-38 / 92±NA[e] / 92.5±NA 39-40 / 95±NA / 97±NA 41 / 96±NA / 106±NA	Gestational age as a continuous variable Adj MDI[f] b = 2.0 (0.45-3.51), p <0.05 Adj PDI[g] b = 3.9 (1.52 - 6.05), p <0.01		Moderate
Rose *et al.* 2013[19] (n = 1562)	Original BSID (Bayley scores of infant development) MDI and PDI	1	GA (n) / MDI mean ± SD / PDI mean (SD) 37 (45) / 102.6±11.4 / 94.4±14.9 38 (260) / 103.4±12.3 / 94.4±14.6 39 (469) / 103.3±12.9 / 96.6±15.4 40 (604) / 105.1±11.5 / 98.5±15.1 41 (184) / 105.4±12.2 / 97.6±14.8	Gestational age as a continuous variable Adj MDI[h] b = 0.8 (0.2-1.4), p = 0.025 Adj PDI[i] b = 1.4 (0.6-2.1), p = 0.036		Moderate
Gyamfi-Bannerman *et al*. 2014[21] (n = 20,093)	Stanford-Binet IQ Full scale intelligence Quotient (IQ) Range 40-160	4	GA (n) / Mean IQ Scores (95%CI) 37 (1290) / 95.9 (93.6 - 95.3) 38 (2361) / 97.6 (95.2 - 96.5) 39 (4040) / 98.6 (97.1 - 98.1) 40 (4816) / 99.8 (98.1 - 99.0) 41 (3782) / 99.8 (99.3 - 100.4) 42 (1953) / 98.1 (99.0 - 100.5)	Test of linear trend P <0.001		High
Yang *et al.* 2009[20] (n = 13,824)	WASI (Wechsler Abbreviated Scale of Intelligence) Full scale intelligence quotient (IQ) Range 40-160	6.5	GA (n) / IQ scores 37 (469) / NA 38 (2100) / NA 39 (11074) / NA 42 (171) / NA 43 (10) / NA	Unadj mean diff (95%CI) -2.6 (-3.7 - -1.4) 0.6 (-1.1- -0.01) REF[k] -1.4 (-3.5 - 0.6) -5.8 (-14 - -2.5)	Adj[j] mean diff (95%CI) -1.7 (-2.7 - -0.7) -0.4 (-1.1 - 0.2) REF -0.4 (-2.5 - 1.7) -5.9 (-15 - 3.3)	Moderate

[b] *GA = Gestational age*

[c] MDI = Mental Developmental Index, *SD = standard deviation*

[d] PDI = Psychomotor Developmental Index

[e] *NA = not available*

[f,g] ***β coefficients*** *represent the actual change in score with each additional week of term,* ***Adj = adjusted*** *for factors ethnicity, parity, maternal age, obstetric risk, birth weight percentile*

[h,i] *Means difference coefficients represent the actual change in score with each additional week of term, Adj = adjusted for factors fetal sex, birth weight centile, socio-economic status*

[k] *REF = Reference category*

[j] *Adjusted for cluster, birth weight, sex, maternal age, maternal height, parental education, and parental occupation*

Table 3 shows the results of the three studies included reporting childhood cognitive outcomes of late preterm birth (34–36 weeks) compared to term birth (37-42 weeks). The abstract by Gyamfi-Bannerman *et al.*(Gyamfi-Bannerman et al. 2014) did not specifically compare late preterm and term deliveries statistically however the results were available for each week of gestation and have therefore been recorded in the table. This was the only study that showed a difference in IQ scores between late preterm (mean IQ 92.5) compared to term born children (mean IQ 98.3) however this was not statistically analysed in the published abstract and standard deviations were not available on contacting the authors. This was a large study (n = 20,093) but was assessed as having a high risk of bias as there was no information on the method of gestational age measurement. The two studies using the Bayley scores of infant development did not show a difference in scores between late preterm and term born infants however this was only done at age 2 and there was no further follow up of the infants. The study by Romeo *et al.*(Romeo et al. 2016) was assessed as having a high risk of bias as there was no mention of how gestational age was measured. The study by Narberhaus *et al.*(Narberhaus et al. 2007) provided the longest follow up of the late preterm-born children, testing IQ using the WISC score (Wechsler Intelligence Scale for Children(Benson 1978)) at ages 13-14. No statistically significant difference was found in the IQ scores between late preterm (mean IQ 112.7, SD [standard deviation] 13.8) and term born children (mean IQ 113.6, SD 11.5). However, these results should be interpreted with caution as the risk of bias was high (no way to determine selective outcome reporting, only some mentioning blinding of outcome assessments and no clear indication of how gestational age was calculated).

Table 9-3 Results of individual studies comparing cognitive outcomes of children born late preterm (34-36 weeks gestation) to term-born infants (37-41 weeks gestation), ordered by participant age at cognitive testing

Study (n)	Cognitive test and overall scores	Age at testing (years)	Main findings Late preterm Mean±SD	Term Mean±SD	Difference in means (95%CI)/significance level	Risk of bias[1]
Nepomnyaschy et al. 2011[22] n = 5,450 (400 late preterm, 5050 term)	BSID-II (Bayley scores of infant development) Bayley short form – mental ability (MDI)[2] Bayley short form – psychomotor ability (PDI) Scores standardised as short form only used (unadjusted range 50-150)	2	MDI 48.9±10 (Range 92.3 – 174.14) PDI 50.2±9.9 (Range 56.43 – 108.53	50.3±10 49.9±10	Unadj MDI -1.43 (-2.70 - -0.16), p <0.05 Adj[3] MDI-0.35 (-1.52 - 0.83), p >0.10 Unadj PDI -0.38 (-1.61 - 0.85), p >0.10 Adj[4] PDI -0.33 (-1.58 - 0.91), p >0.10	Moderate
Romeo et al. 2015[23] n = 119 (71 late preterm, 48 term)	BSID-II (Bayley scores of infant development) MDI only (range 50-150)	2	96.7±9.3	97.1±6.5	p > 0.05	High
Gyamfi-Bannerman et al. 2014[21] n = 20,093 (1,951 late preterm, 18,242 term)	Stanford-Binet IQ+ Full scale intelligence Quotient (IQ) Range 40-160	4	92.5±NA[5]	98.3±NA	NA	High
Naberhaus et al. 2007[24] n = 64 (11 late preterm, 53 term)	WISC (Wechsler Intelligence Scale for Children) Full scale intelligence quotient Range 40-160	13-14	112.7±13.8	113.6±11.5	p > 0.05	High

[1] *Risk of bias is a summary using the RoBANs tool, full details in appendix 3*

[2] MDI = Mental Developmental Index, PDI = Psychomotor Developmental Index

[3] *Coefficient represent the actual change in score associated with being late-preterm versus full term and Adj. = adjusted for race/ethnicity, age, education, marital birth, father co-residence, household residence, household below poverty*

[4] *Coefficient represents the actual change in score associated with being late-preterm versus full term and Adj. = adjusted for race/ethnicity, age, education, marital birth, father co-residence, household residence, household below poverty*

[5] *NA = not available*

9.4 Discussion

9.4.1 Main Findings

In this systematic review of seven studies (reporting on 41,433 children), the four studies investigating IQ scores within term deliveries found that children born at early term (37-38 weeks) had lower IQ scores at ages one, four and six compared to those born at full term (39-41 weeks). One study (n = 13,824)(Yang et al. 2010) found a decrease in IQ score at >42 weeks. In the four studies comparing late-preterm deliveries (34-36 weeks gestation) to their term counterparts there were no differences in cognitive outcomes at ages two, four and 14. Studies were heterogeneous, and several were at high risk of bias and therefore summary effect sizes cannot be reported. No studies were identified comparing outcomes between the ages of four and 14. However, it is useful to consider individual study effect sizes and the potential effect on clinical practice. For example, a three point difference in the Stanford-Binet IQ test between children born at 37 weeks and those born at 41 weeks (Gyamfi-Bannerman et al. 2014) may not be important at an individual level but this can have important implications at a population level and should be considered along with other factors (estimated birthweight, obstetric risk factors) when clinicians are discussing timing of delivery with parents.

9.4.2 Strengths and limitations

The strengths of this review include the comprehensive and extensive search strategy, with no language restriction, combined with a detailed pre-defined eligibility criteria for study selection. At the screening stage, to reduce reporter bias, two reviewers independently screened a selected sample to check for accuracy and agreement regarding inclusion of studies. Two reviewers critically appraised all included studies independently. A wide range of cognitive assessments was used in the included

studies providing a good overview of the various tests available, but this does make comparison between studies more difficult.

Despite the comprehensive nature of the search, the possibility of missing relevant papers cannot be excluded. We did not have the resources to translate the papers in foreign languages however we did non-expertly translate to see if any papers fitted the inclusion criteria and none were thought to be relevant. Another limitation was the problems encountered with categorisations of gestational age. A number of studies (22 studies reporting cognitive outcomes of 3,357 infants) only listed <37 weeks of gestational age (all preterm births) which inevitably included those <34 weeks and therefore the whole study was excluded. This may have potentially exacerbated the risk of publication bias as we excluded these studies. Due to limited resources, attempts to contact the individual authors of these studies to see if data was available for 34-36 weeks of gestation was not performed. At delivery, birthweight and gestational age are highly correlated. There is a small but statistically significant correlation between birthweight and cognition in childhood and adulthood (each 1kg increase is associated with 0.13 standard deviation test score increase)(Grove et al. 2017) Some studies did account for birthweight in analyses, and some did not. This may not be appropriate due to their high correlation, and birth weight may be a mediator of the relationship between gestational age and cognition, rather than a confounder. Future studies should report both birth weight and gestational age as a continuous measure, allow their relative contributions to be measured. Structural equation modelling or similar techniques could be used to model the potential competing causal pathways. We excluded studies with intra-uterine growth restriction (IUGR) because we wanted to study normal healthy singletons, appropriate for gestational age and IUGR may be associated with adverse cognitive outcomes. This review is based on observational data with high levels of between-study heterogeneity and therefore statistical analysis of the studies was not possible given that the studies were not directly comparable. Limited conclusions can be made regarding the mechanism of action of gestational age on long-term cognitive outcomes because of the nature of the observational data. There were a number of

potential sources of bias across the included studies. Although most studies stated how the participants in the studies were chosen in an attempt to reduce selection bias, it is difficult to determine generalizability of the results outwith the populations that were studied. There was a large variation in the number of confounders (table 1) adjusted for the in various studies and many did not account for indication for delivery and some did not account for socio-economic factors (strongly associated with cognitive outcomes) and therefore there is a risk of residual confounding among the studies.

9.4.3 Interpretation

Comparing the outcomes within the weeks of term (37-42 weeks) this review has shown that cognitive scores in childhood differ throughout the weeks of term delivery and are lowest in those individuals born in early term gestation (37-38 weeks) when measured at ages one, four and six. Although this review specifically set out to look at cognitive outcomes, the findings are in line with those studies of school performance of individuals born within the different weeks of term. Two large population based studies (Figlio et al. 2016, Smithers et al. 2015) have recently published school outcomes of individuals born within the weeks of term. Smithers *et al.*(Smithers et al. 2015)(n = 12,601) showed that children born at 40-41 weeks gestation had the lowest risk of vulnerability at school aged five compared to those born early term (37-39 weeks) or post term (42-45 weeks). Figlio *et al.*(Figlio et al. 2016)(n = 1,536,482) showed that children born late term (41 weeks) performed better in school at the age of five through to 18 compared to those born at full term (39 or 40 weeks). Only one of the studies in the review looked at the effect of post-term delivery (>42 weeks) on cognitive outcomes and found those individuals to have a lower score compared to full term (39-40 weeks). This U-shaped relationship has previously been described in the study by MacKay *et al.* (n = 407,503) which found the lowest risk of special educational need at school in those born at 41 weeks gestation compared to those born <41 weeks and >41 weeks (MacKay et al. 2010). We identified a previous systematic review published in 2015 which also found a

reduction in long term cognitive outcomes of children born early term compared to those born full term but we were unable to reconcile data included in this previous review with source data (Chan et al. 2016).

The mechanism of early term (37-38 week) delivery leading to lower cognitive outcome scores compared to full term deliveries (39-41 weeks) is likely to be multifactorial. Vohr *et al.* described how brain weight increases rapidly in the last trimester of pregnancy with brain weight at 38 weeks 90% of the weight at full term which may account for the increased vulnerability of early term infants at school (Vohr 2013). For those born post term (>42 weeks) it is thought the increased vulnerability at school age is due to poorer placental perfusion (Link, Clark and Lang 2007).

Only three studies were included in the review comparing the cognitive outcomes of children born late preterm (34-36 weeks) with those born at term (37-41 weeks). There were no (statistically or clinically) significant differences in cognition found between these groups at ages two, four and 13. However, the quality of evidence from these observational studies is poor due to high risk of bias (high chance of residual confounding, no outcome assessor blinding and no way to ascertain if selective outcome reporting took place). We therefore do not make any clinical recommendations relating to the timing of delivery, as these observational data cannot be used for this purpose. The included studies all used term deliveries as the control group, but as this includes early term (which, as described above, have lower cognitive outcomes than 39+ weeks) there is a possibility that differences between <37 weeks and later were masked. The conference abstract, although it did not specifically compare the results for late preterm versus term delivery if we compared late preterm deliveries (mean IQ 92.5) with only full term deliveries (39-41 weeks, mean IQ 98.3) the difference is large and provides further evidence of a partial dilution of the results in treating term deliveries as a continuum. This was a large study however the data was taken from an old cohort study performed between 1959-1966 and many of the variables, including method of gestational age measurement were not available. 3 out

of 4 of the studies had a high risk of bias, and they all assumed homogeneity between the term cases, which, as shown above, is not the case. Although a previous systematic review has shown a clear increase in physical morbidity associated with late preterm delivery (Teune et al. 2011)(34-36 weeks) compared to 37+ weeks, there remains a paucity of evidence regarding long term cognitive outcomes in this group. Future studies should use a full-term delivery group (39-41 weeks) as the control group and adopt uniform gestational age categorizations, ideally with similar outcome measures, to allow for easy comparison between studies. Individual level data should be made available as soon as possible to allow large scale individual participant meta-analysis.

9.5 Conclusion

Overall this systematic review has found that children born at full term (39-41 weeks) have the highest cognitive outcome scores compared to those born at early term (37-38 weeks). Given the high prevalence of early term deliveries (the fastest growing proportion of singleton births in the US), small differences at an individual level in cognitive outcomes are likely to have a large impact at a population level. Further research is required to look at the potential reasons for this, and to consider outcomes of late-preterm delivery using a suitable control group of full term (39-41 weeks). The findings from this review have important implications for clinicians and the long-term cognitive outcomes based on gestation at delivery should be discussed with parents regarding optimum timing of delivery.

9.6 Appendices

9.6.1 Appendix 1: Search Strategies (Ovid MEDLINE/EMBASE (R) In-Process & Other Non-Indexed Citations (1946 to Present), Searched on: 5 August 2016)

Search	Results
1	gestational age/ or premature birth/ or preterm birth/ or preterm infant/ or premature infant/
2	(gestation* adj3 age*).tw.
3	((pre-term* adj3 birth*) or (preterm* adj3 birth*) or (pre-term* adj3 born*) or (preterm* adj3 born*)).tw.
4	((prematur* adj3 born*) or (prematur* adj3 birth*) or (pre-matur* adj3 born*) or (pre-matur* adj3 birth*)).tw.
5	((prematur* adj3 bab*) or (prematur* adj3 infan*) or (pre-matur* adj3 infan*)).tw.
6	((preterm* adj3 bab*) or (preterm* adj3 infan*) or (pre-term* adj3 bab*) or (pre-term* adj3 infan*)).tw.
7	1 or 2 or 3 or 4 or 5 or 6
8	((moderate* adj5 preterm*) or (moderate* adj5 pre-term*) or (moderate* adj5 prematur*) or (moderate* adj5 birth*) or (moderate* adj5 born*)).tw.
9	((late adj5 preterm*) or (late adj5 pre-term*) or (late adj5 prematur*)).tw.
10	((near term adj5 birth*) or (near term adj5 born*) or (near term adj5 infan*) or (near term adj5 bab*)).tw.
11	early term.tw.
12	8 or 9 or 10 or 11
13	7 or 12
14	(preterm* labo?r or pre-term* labo?r or (prematur* adj2 labo?r) or (pre-matur* adj2 labo?r)).tw.
15	((preterm* adj2 deliver*) or (pre-term* adj2 deliver*) or (prematur* adj2 deliver*) or (pre-matur* adj2 deliver*)).tw.
16	(wom#n or matern* or pregnan*).tw.
17	(resus* or (infan* adj3 dea*) or (infan* adj3 mort*) or ((neonat* adj3 dea*) or (neonat* adj3 mort*))).tw.
18	(syndrom* or palsy or congenital* or deform*).tw.
19	(feed* or nutri*).tw.
20	(heart* or cardio* or lung* or hypox* or pulmon*).tw.
21	14 or 15 or 16 or 17 or 18 or 19 or 20
22	13 not 21
23	cognition/ or intelligence/ or educational status/ or child development/
24	exp mental process$/ or exp aptitude tests/ or exp neuropsychological tests/ or exp psychometrics/ or exp educational measurement/
25	(cogniti* or education* or intelligen* or IQ).tw.
26	(((aptitude or neuropsycholog*) adj5 test*) or ((aptitude or neuropsycholog*) adj5 assess*)).tw.
27	((adolescen* adj10 cogniti*) or (teenage* adj10 cogniti*)).tw.
28	23 or 24 or 25 or 26 or 27
29	22 and 28
30	(nurs* or (care adj3 quali*) or (care adj3 facilit*) or (hospital* adj facilit*) or (hospital* adj care*)).tw.
31	29 not 30
32	**remove duplicates from 31**

9.6.2 Appendix 2: Eligibility criteria for study selection

	Inclusion criteria	Exclusion criteria
Population	Normal healthy singletons, appropriate-for-gestational-age, without developmental delay. Low birth weight, but appropriate-for-gestational-age could be included. May include studies investigating the relationship between brain imaging/ neurological tests and cognitive assessment scales on **a group of healthy singletons**.	▪ Multiple births/ twins ▪ Infants with congenital or chromosomal abnormality ▪ Small for gestational age/ IUGR ▪ Neuro-developmental delay ▪ Cognitive impairment ▪ Learning disability ▪ Psychiatric illness ▪ Effects of medication/drugs/infection on cognition ▪ Interventions used to improve cognitive outcomes (e.g. Infant Health & Development Program (**IHDP**); Mother-Infant Transaction Program (**MITP**); Infant Behavioural Assessment and Intervention Program (**IBAIP**); Newborn Individualized Developmental Care and Assessment Program (**NIDCAP**); massage therapy, kangaroo care, breastfeeding, DHA supplementation, etc) ▪ Studies that use of healthy preterm/term singletons as a control group for comparison with another group of preterm infants with identified disease (e.g. lung injury, brain injury, intraventricular or periventricular haemorrhage, heart disease, sepsis) in terms of: - efficacy of brain screening tools, and - NICU or post-natal intervention.
Main subject of interest	Studies should have gestational age as their primary aim. Gestational age can be measured by last menstrual period, or ultrasound, or both.	Birth weight only, without any mention of gestational age.
Methods to measure cognition	Any tests of cognitive abilities that assess verbal reasoning, plus numeric processing (or problem solving) skills. Examples of tests: ▪ British ability scales ▪ Woodcock-Johnson Tests of cognitive abilities ▪ Stanford-Binet test ▪ Bayley scales of infant development ▪ Wechsler intelligence scale for children ▪ Wechsler preschool & primary scale of intelligence ▪ Kaufmann-ABC ▪ Terman-Merrill ▪ Ravens Test/ Ravens progressive matrices ▪ Uzgiris-Hunt Scales ▪ McCarthy Scales of Children's Abilities ▪ Revised Amsterdam Child Intelligence test ▪ Miller Assessment for Preschoolers	▪ School examination results or school performance ▪ **Parent-reported** questionnaires on child behaviour and development (e.g. Ages and Stages Developmental Questionnaire) ▪ **Neurodevelopmental tests:** - focussing only on gross motor skills/ function o Bender-Gesalt Motor Test o Pretchl's General Movement Assessment o Movement-ABC o Movement Assessment of Infants (MAI) - screening for neurodevelopmental delay: o Baley Infant Neurodevelopmental Screen o Cognitive Adaptive Test/ CLAMS o Denver Developmental Screening Test o Gesell Developmental Scale o Kyoto Scale – screen for pervasive developmental disorder o First STEP screening test for preschoolers ▪ **Psychomotor/ neuromotor tests** o Brunet-Lezine Scale o Baley Scale of Infant Development, using only Psychomotor Developmental Index subscale ▪ **Behavioural tests** o Brazelton Scale o Assessment of preterm infant behaviour (APIB) o Neurobehavioural Assessment of the Preterm

9.6.3 Appendix 3: study quality assessed using the RoBANs* tool

Study	Bias in selection of participants	Bias due to confounding	Bias in measurement of interventions	Blinding of outcome assessments	Incomplete outcome data	Selective outcome reporting	Overall judgement
Comparison within term (37-42 weeks)							
Espel et al. 2014. (USA)	Low Longitudinal prospective study	Low Adjusted for multiple confounders	Low GA measured by USS. Bayley Scales used for IQ with reviewed examiners.	High Not mentioned	High Not mentioned	Unclear	Moderate
Rose et al.2013 (Chile)	Low Taken from RCT study sample	Low Adjusted for multiple appropriate confounders	Low Taken from mother-reported last menstrual period only	High No mention of blinding	High Missing data not reported	Unclear	Moderate
Gyamfi Bannerman et al. 2014 (USA)	Low Taken from prospective cohort study 1959-1966	Low Adjusted for multiple appropriate confounders	High No mention of how GA measured and note old sample	High Not recorded	Unclear	Unclear	High
Yang et al. 2009. (Belarus)	Low Subjects taken from the PROBIT RCT. Samples representative of the whole population	Low Adjusted for multiple appropriate confounders	Low Based on hospital USS	High No mention of blinding	Low Missing outcome data reported	Unclear	Moderate
Late preterm vs term (34 – 36+6 weeks)							
Nepomyanschy *et al.* 2011 (USA)	Low Convenience Sample taken for ECLS (nationally representative study sample)	Low Adjusted for multiple appropriate confounders	HIgh Gestational age taken from mother-reported last menstrual period, LMP not available in 250 (0.5%) women and estimated from birth certificate	Low Interviewer blinded	Low No mention of incomplete outcome data	Unclear	Moderate
Romeo *et al*. 2015 (Italy)	Unclear Cases/controls taken from one hospital cohort however control group not matched.	High No mention of confounders adjusted for in multivariate analysis other than gender	High No mention of how GA was measured	Low Blinding of assessors	Low Non-attenders were excluded from the results	Unclear	High
Narberhaus *et al*. 2007 (Spain)	Low Matched case control with both samples from the same clinic.	Low Adjusted for multiple appropriate confounders	High No statement of how GA was measured	High No mention of blinding	High No mention of incomplete outcome data	Unclear	High

9.6.4 Appendix 4: Description of the cognitive tests included in the review

Measure	Description and Sources
Bayley scores of infant development (BSID)	• Original Bayley scores – monitor neurodevelopmental outcomes up to the age of 3. This is split into: - The Mental development index (MDI) which measures distinct cognitive, receptive lauguage and expressive language scales and; - The Psychomotor development index (PDI) which measures fine and gross motor skills • BSID-II is the revised bayley scores which was published in 1993 • The Bayley short form includes a subset of score from BSID-II which includess some assessments from both the MDI and the PDI • Raw scores are converted into standard scores based on the the child's chronological age (mean 100, SD 15 with a range from 50-150) • Scores are interpreted as below: ⩽ 69 = significantly delayed performance 70-84 = mildly delayed performance 85-114 = within normal limits ≥ 115 = accelerated performance
WISC (Weschler Intelligence Scale for Children) Full scale intelligence quotient	• Intelligence test for children between the ages of 6 and 16. • A Full scale IQ should be generated representing a child's general intellectual ability. It provides five primary index scores (verbal comprehension index, visual spatial indexm fluid reasoning index, working memory index, and processing speed index) • Scores follow that of general IQ scores (<70 – Impaired or delayed, 70-79 – borderline, 80-89 – low average, 90-110 – average, 110-119- high average, 120-129 – above average, >130 – gifted/superior)
WASI (Wechsler Abbreviated Scale of Intelligence) Full scale intelligence quotient (IQ)	• IQ test designed to measure intelligence and cognitive ability in adults and older adolescents • Most commonly used intelligence test • Scores follow that of general IQ scores (<70 – Impaired or delayed, 70-79 – borderline, 80-89 – low average, 90-110 – average, 110-119- high average, 120-129 – above average, >130 – gifted/superior)
Stanford-Binet IQ Full scale intelligence quotient (IQ)	• Cognitive ability and intelligence test • Can be used from the age of two • Scores follow that of general IQ scores (<70 – Impaired or delayed, 70-79 – borderline, 80-89 – low average, 90-110 – average, 110-119- high average, 120-129 – above average, >130 – gifted/superior)

9.6.5 Appendix 5: PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	5
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	7
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	8
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	9
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	9/10/ App 2
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	9/10
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	App 1

Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	9/10/App 2
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	10
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	9/10
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	10/App 3
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	10
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	10
Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	10/App 3
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	10
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	10/11/Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	12/13/App 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 2/3

Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	App 3
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	13/14
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	14/15
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15/16/17
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	19

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

9.6 Chapter Conclusion

The work presented in **Chapter 9** demonstrated that children born at full term (39-41 weeks) have higher cognitive scores than children born at early term (37-38 weeks). No difference was found in cognitive scores between those children born late preterm (34-36 weeks) compared to those born at term (37-43), however the quality of these studies was noted to be poor and the range across the whole spectrum of term was large which may have masked any differences.

Chapter 9 together with **Chapter 8** highlights the need to consider the long-term outcomes as well as short-term outcomes of gestation at delivery in both twin and singleton pregnancies. This information should be used by healthcare professionals providing antenatal care and by women and families when making decisions regarding timing of delivery.

Chapter 10

Discussion

10.1 Summary of Results

The work presented in this thesis was based on the hypothesis that timing of birth impacts on both short- and long-term offspring outcomes. This was investigated in singletons by carrying out a systematic review of long-term cognitive outcomes according to gestation at birth and a population cohort study investigating environmental influences on preterm birth. In twins, timing of birth was investigated in the form of two population-based cohort studies, one of which included record linkage of maternity data to education data providing short- and long-term follow up of the twin offspring. Outcomes in twins were then compared to outcomes in singletons in the form of a population cohort study.

The prevalence of twin pregnancy has been increasing in recent years (mainly due to assisted reproduction technologies) and it is therefore becoming increasingly important to consider outcomes in the twin offspring. Optimising the timing of birth is a key strategy in minimising fetal death and other adverse outcomes. Evidence supporting the gestation at delivery is needed to guide clinicians regarding the management of twin pregnancies.

This thesis showed that timing of birth is important for short- and long-term childhood outcomes in both singleton and twin pregnancy. In twins, the optimum timing of delivery of uncomplicated twin pregnancy is 37 weeks and this result is of great importance for clinicians in the management of twin pregnancy.

In **Chapter 5**, it was shown that, after adjustment for known risk factors, large geographical differences in rates of preterm birth remain within countries. Using Sweden as a model of a very high human-development index country, with a homogeneous population and comprehensive healthcare system, why such extreme geographical differences exist is still unknown. In order to investigate this further an exploratory analysis was performed to investigate rural and urban environmental and socioeconomic factors on preterm birth rates. Longer gestational length was found in urban areas compared to rural areas forming the hypothesis that the underlying mechanism is access to tertiary healthcare facilities.

In **Chapter 6**, it was shown that delivery of twins at 37-38 weeks is associated with the lowest risk of perinatal death. There was a 2-fold increase in perinatal death with delivery before 37 weeks and at or beyond 39 weeks, after adjustment for known potential confounders. These findings were in keeping with the current UK NICE guidance on delivery of dichorionic twins at 37-38 weeks to reduce the risk of perinatal death. A 2-fold increase in perinatal death was found in monochorionic twins compared to dichorionic twins. It was not possible to determine a difference in perinatal death in individual weeks of gestation in dichorionic and monochorionic twins due to the small sample size in each week of gestation. The outcomes in twins conceived through ART were the same as the naturally conceived twins with no increased risk of preterm delivery or perinatal death. Twins conceived through ART should, therefore, be managed according to the current UK guidance and this information can be used by clinicians and parents when planning antenatal clinical management and advising families with a twin pregnancy.

In **Chapter 7**, using a population study of routinely collected Scottish maternity data linked to childhood educational data, the optimal week of gestation for delivery of uncomplicated twin pregnancy in terms of both short- and long-term offspring outcomes was 37 weeks. Being born before 37 weeks was associated with an increased risk of both perinatal mortality and having a record of SEN at school. The risk of stillbirth and neonatal death were balanced at 37 weeks and the risk of SEN

did not increase for births beyond 37 weeks gestation. These data are in keeping with the current clinical guidance in the UK and the information should be considered by women expecting twins and the clinicians managing and making decisions regarding their timing of birth.

In **Chapter 8**, using a population cohort study design to determine the differences in short-term perinatal outcomes between twins and singletons, it was shown that overall twins had higher odds of stillbirth and NND. The odds of stillbirth were higher in twins across the full range of gestation compared to singletons. The odds of NND were higher in twins in the extreme preterm period (less than 28 weeks) compared to singletons. The odds of NND were then lower in twins from 28-37 weeks compared to singletons. It is hypothesised this may be due to the different aetiology of preterm birth in twins compared to singletons with preterm twins born in the context of uterine stretch compared to preterm singletons born in the context of pathological processes such as pre-eclampsia or infection. The study also highlighted the importance of using the correct denominator as the results were very different to previous studies which used livebirths as the denominator for stillbirth compared to ongoing pregnancies at risk presented in this chapter. Clinically, the results presented in this chapter are more plausible than previous studies reporting a lower risk of perinatal death in preterm twins compared to preterm singletons.

In **Chapter 9**, a systematic review of seven studies reporting on 41,344 children found higher cognitive scores in infants born at full term (39-41 weeks) compared to those born at early term (37-38 weeks) with progressively higher scores at each week of term. The four studies comparing long-term cognitive outcomes of late preterm births (34-36 weeks) to 'term' births (37-42 weeks) found no difference in cognitive scores between the two groups, however the data were scarce. A narrative review only was undertaken due to the different cognitive tests and scores reported and highlighted the need for future studies to adopt a standard gestational age control group (ideally 39-41 weeks) to allow meaningful comparisons between studies to be drawn.

10.2 Wider and Clinical Implications

10.2.1 Timing of Birth of Singletons

Gestational age at birth has been steadily decreasing in over recent years with most singletons now being delivered during the 39th or 40th week (Ananth et al. 2018). The national policy in the UK is to offer routine IOL from 41 weeks gestation in singleton pregnancies. Evidence supporting earlier IOL for all women (i.e at 39 rather than 41 weeks) is growing with a recent large meta-analysis of IOL at 39 weeks compared to expectant management reporting a significant reduction in caesarean section rates, peripartum infection, meconium aspiration syndrome, respiratory morbidity in the neonatal and perinatal mortality in the IOL group compared to the expectantly managed group (Grobman and Caughey 2019). Although it was concluded that the results should reassure women and care providers that IOL at 39 weeks is a reasonable choice in terms of obstetrical outcomes it highlighted the need for obstetric units to find new ways to accommodate larger numbers of women with longer lengths of stay in hospital (Greene 2018). A planned secondary health economic analysis of the RCT of IOL at 39 weeks versus expectant management found that although women undergoing IOL spent more time in the labour and delivery suite, they had fewer antepartum visits, tests and intrapartum interventions and shorter postpartum maternal and neonatal hospital stays (Grobman 2019). Early IOL appears to be a reasonable option for women with an uncomplicated singleton pregnancy in terms of short-term perinatal outcomes and resource utilization but it is important to consider the long-term offspring outcomes associated with earlier delivery.

The findings of the systematic review presented in this thesis suggest that the long-term follow up of the offspring is important to consider when determining optimum gestation at delivery. The review found higher cognitive scores in later life in offspring who delivered at full-term (39-41 weeks) compared to those who delivered at early term (37-38 weeks). This is in keeping with a previous population cohort

study looking at the risk of having SEN at school according to gestation at delivery which found this risk to be lowest at 41 weeks (MacKay et al. 2010).

10.2.2 Preterm Birth in Singletons

Preterm birth remains a major cause worldwide of perinatal morbidity and mortality in singleton pregnancies. One of the key problems with preterm birth is that the aetiology remains unclear and therefore guiding effecting interventions and treatments is difficult if the underlying mechanism is not clear. A recent review highlighted the need for further investigation into the different mechanisms driving preterm birth in singletons (Stock and Ismail 2016). The findings presented in this thesis suggest that geographical and environmental factors do have a role to play in preterm birth in singletons, but this area requires further research. The finding that pregnancies in urban areas have longer gestational lengths is novel and will guide future research to explore further this association. The accessibility of specialised obstetric care providers in Sweden is now being mapped to preterm birth rates to see if this is potentially the underlying factor driving the differences in preterm birth rates across Sweden.

10.2.3 Timing of Birth of Twins

To the best of our knowledge, the work in this thesis is the first work to link maternity records to education records for a population of twin pregnancies in the UK to provide long-term follow up. The UK NICE Multiple Pregnancy Guideline Development group have identified the following research priority for twin pregnancy: further research is necessary to investigate perinatal and neonatal morbidity and mortality in babies born by elective birth in twin pregnancies; the work presented in this thesis aimed to address this research priority (Visintin et al. 2011). The current UK guidance regarding timing of delivery of twins is based on evidence from two population studies of fetal death (Kahn et al. 2003, Minakami and Sato 1996), neither of which provide any long-term follow up of the twin infants. The findings presented in this

thesis suggest that birth at 37 weeks gestation is associated with optimal short- and long-term outcomes for twin babies. This is in keeping with the current guidance and the most recent systematic review on timing of delivery (Cheong-See et al. 2016) but for the first time provides long-term offspring follow up. The results provide reassurance to clinicians and women pregnant with twins that the current policy is not associated with excess harm and that delivery from 37 weeks gestation optimises long-term outcomes in the twin infants too. The study also provides clear evidence that delivery of uncomplicated twins before 37 weeks is associated with increased risks of both perinatal death and SEN at school.

10.2.4 Perinatal Outcomes in Twins Compared to Singletons

Whilst it has long been established that twins have an increased risk of perinatal mortality compared to singletons (Manktelow et al. 2014), the evidence examining the differences in perinatal mortality in twins and singletons born preterm remains inconclusive. Previous studies found a lower perinatal mortality rate in twins born preterm (<37 weeks) compared to singletons born preterm (Vasak et al. 2017, Minakami and Sato 1996). This is a surprising result and it is likely that these previous studies used flawed methodology because of how the numerators and denominators were derived. The information regarding the differences in perinatal outcomes between twins and singletons born preterm is required however to guide clinical decision making and accurate counselling of parents of twins. The results presented in this thesis demonstrated that, contrary to the results presented in previous studies, twins born at all gestational ages had a greater risk of stillbirth compared to singletons. Interestingly twins had lower rates of NND compared to singletons in the gestational period of 29-37 weeks and this likely reflects the different aetiology of preterm delivery in twins. This study confirms the need for increased monitoring and antenatal care of twin pregnancies due the increased risks of stillbirth across the range of gestation. It also highlights that although the aetiology of twin infants could be a biologically plausible reason for the difference in NND rates between twins and singletons, differences in neonatal care for twins compared to singletons should also

be considered as a potential reason for the lower NND rates in twins and this requires further research.

10.3 Strengths and Limitations

The strengths of the studies presented in this thesis are the large sample sizes and unselected study populations, therefore maximising study power and reducing the risk of selection bias. The other main strength is the novel record linkage of Scottish twins to their education data providing an opportunity to study long-term outcomes using a cohort design. This has not been attempted previously in the Scottish data because of challenges in doing so: the twins have the same postcode, maternal ID and date of birth and so differentiating between them is difficult. We performed the education data linkage on sex discordant twins only to ensure the correct twin was linked with their education data and received a high match rate of 98%.

In **Chapters 7** and **8**, the most accurate denominators were used to identify the pregnancies at risk. GEE analysis was also employed to adjust for the potential clustering effect of the twin infants. Previous studies were flawed by failing to address both of these issues.

The work presented in this thesis highlighted the need to consider short- and long-term outcomes when determining optimum timing of birth. In **Chapter 9**, the systematic review aimed to address the paucity of evidence on long-term cognitive outcomes of singletons according to gestation at delivery using a comprehensive and extensive search strategy and pre-defined eligibility criteria. Again, the large sample size of participants in this systematic review (41,344 children) was one of the main strengths of the study. However, the review also highlighted that the heterogeneity of the studies included in the review made it impossible to synthesise the results in a meaningful meta-analysis.

Limitations of the studies presented in this thesis mainly involve general issues regarding routinely collected data and are covered in the relevant data chapters. One of the main steps in the data cleaning process described in **Chapter 4** is data reduction. Although this is a necessary process to prepare the variables for inclusion in the multivariate analyses and allow easy interpretation of the results, grouping of numerical variables does lead to a loss of information (Kirkwood and Sterne 2010). It is therefore very important to categorise the variables based on clinical relevance or to be consistent with previous studies and this was performed in **Chapters 5-8** when selecting potential confounding variables.

The methods presented in this thesis are a strength as outlined above, but also have some limitations. In **Chapter 6,** a Cox regression model was used to determine the association between gestational week of delivery and the risk of perinatal death and the results were presented as adjusted HRs for each week of gestation. Gestation was treated as a continuous variable with 37 weeks set as the referent. Although the results were easy to interpret and model via a Kaplan-Meier survival analysis, the method failed to accurately determine the population at risk as it compared death at each week to the referent of 37 weeks as opposed to ongoing pregnancies at risk. Another limitation of the methods is the competing risk analysis employed in **Chapter 7**. A simple numerical method of risk versus benefit was presented, as described in previous studies of twins (Cheong-See et al. 2016), concluding that the risks of stillbirth and neonatal death were balanced at 37 weeks. The limitation of this method is the failure to adjust the results for potential confounders. For example, the Fine-Gray model of competing risk describes the relative effects of covariates on the rate at which events occur (Austin and Fine 2017). Had this method been employed, an estimate adjusted for potential confounders could have been presented.

Another potential limitation is the handling of missing data employed in this thesis. Multiple imputation was a skill gained in the final year of the PhD and was therefore only used for the analysis in Chapter 7.

10.4 Future Directions

Targeted interventions to prevent preterm birth in singletons are limited. This is due to the combination of a lack of specific targets for prevention due to the wide aetiology of preterm birth and also the unknown implications of implementing these interventions in pregnancy. Given that preterm birth has many implications in later life for the offspring including and increased risk of long-term neurological disability, there is an urgent need for further targets for prevention of preterm labour and interventions to reduce the risk following preterm birth. The work presented in this thesis suggests that environmental and geographical factors likely have a role to play in preterm birth and provides a basis for a number of avenues that can now be explored in more detail. Two recent studies have shown that high levels of sunlight (and therefore vitamin D) are associated with a reduction in preterm birth rates (De-Regil et al. 2016) but high levels of air pollution associated with and increased risk of preterm birth (Hao et al. 2015), using Sweden as a model of a country with a very high human development index the effect of these environmental influences on a whole country population could be explored.

Gestational age at delivery has decreased over the last decade with emerging evidence that IOL at 39 weeks is a reasonable choice in term of short-term offspring outcomes, even in uncomplicated singleton pregnancies (Grobman et al. 2018, Grobman and Caughey 2019). Work presented in this thesis suggests that, in terms of long-term offspring outcomes, IOL at 39 weeks is a reasonable choice for singletons but delivery before 39 weeks (37-38 weeks) may be associated with decreased cognitive outcomes in the offspring. The systematic review highlighted the difficulty with meta-analysis due to different methods of cognitive outcome scoring and the need for further research with standard gestational age categories to allow meta-analysis of studies. Previous cohort studies have shown that the risk of SEN in singletons, is in fact lowest at 41 weeks of gestation (MacKay et al. 2010). The record linkage described in this thesis between maternity records and the long-term education records in the offspring could be used in future studies to investigate the long-term educational outcomes in

offspring who have been induced compared to those expectantly managed in order to determine if IOL at 39 weeks is a reasonable option for uncomplicated singleton pregnancies in terms of long-term offspring outcomes.

Studies reporting on longer-term twin offspring outcomes are limited. The work presented in this thesis is the first record linkage of maternity record to education records in twins in the UK. Building on this novel data linkage further research priorities could be investigated in twins. For example, it is well established that second twins born at term have increased risk of perinatal morbidity and mortality compared to first twins (Smith et al. 2002, Smith, Fleming and White 2007), but the long-term offspring outcomes of second twins compared to first twins has not been investigated. The novel twin record linkage described in this thesis could be used in future studies to determine if the increased risk of short-term adverse outcome in second twins also continues into later life. This would be useful information for counselling parents of twins and potentially putting into place additional educational measures if they are known to be at increased risk.

Preterm birth is the most common adverse perinatal outcome in twin pregnancy, with half of all twins delivering less than 37 weeks (ISD Scotland 2009) and this results in short- and long-term morbidity in the twin offspring. The work demonstrated in this thesis demonstrated that twins were at increased risk of adverse perinatal outcome compared to singletons at every week of preterm gestation. Future studies could include more detailed analysis of the aetiology of preterm birth in twins, similar to work presented in this thesis in singletons. A better understanding of the aetiology of preterm birth in twins could potentially guide interventions to prolong gestation in twins at risk of preterm birth.

References

Aagaard, K., K. Riehle, T.-A. Mistretta, C. Coarfa, S. Raza, S. Rosenbaum, I. Van den Veyver, A. Milosavljevic, D. Gevers, J. Petrosino, J. Versalovic, J. Ma, C. Huttenhower & N. Segata (2012) A Metagenomic Approach to Characterization of the Vaginal Microbiome Signature in Pregnancy. *PLoS ONE,* 7(6):e36466.

ACOG committee opinion no. 559 (2013): cesarean delivery on maternal request. *Obstetrics and gynecology,* 121**,** 904.

Alfirevic, Z., T. Stampalija, D. Roberts & A. L. Jorgensen (2012) Cervical stitch (cerclage) for preventing preterm birth in singleton pregnancy. *Cochrane database of systematic reviews (Online),* 4**,** CD008991.

Allison, P. (2012) Why you probably need more imputations than you think. *Statistical Horizons*. https://statisticalhorizons.com/more-imputations

Ananth, C. V., R. L. Goldenberg, A. M. Friedman & A. M. Vintzileos (2018) Association of temporal changes in gestational age with perinatal mortality in the United States, 2007-2015. *JAMA pediatrics,* 172**,** 627-634.

Arabin, B. & Z. Alfirevic (2013) Cervical pessaries for prevention of spontaneous preterm birth: past, present and future. *Ultrasound in Obstetrics & Gynecology,* 42**,** 390-399.

Austin, P. C. & J. P. Fine (2017) Practical recommendations for reporting F ine-G ray model analyses for competing risk data. *Statistics in medicine,* 36**,** 4391-4400.

Ayorinde, A. A., K. Wilde, J. Lemon, D. Campbell & S. Bhattacharya (2016) Data resource profile: the Aberdeen maternity and neonatal databank (AMND). *International journal of epidemiology,* 45**,** 389-394.

Babatunde, O. A., Adebamowo S.N., Ajayi I.O & Adebamowo C.A. (2018) Neurodevelopmental Outcomes of Twins Compared With Singleton Children: A Systematic Review. *Twin Res Hum Genet* 21**,** 136-145.

Barrett, J. F. R., M. E. Hannah, E. K. Hutton, A. R. Willan, A. C. Allen, B. A. Armson, A. Gafni, K. S. Joseph, D. Mason, A. Ohlsson, S. Ross, J. J. Sanchez & E. V. Asztalos (2013) A Randomized Trial of Planned Cesarean or Vaginal Delivery for Twin Pregnancy. *The New England Journal of Medicine,* 369**,** 1295-1305.

Bauer, J., D. A. Groneberg, W. Maier, R. Manek, F. Louwen & D. Brüggmann (2017) Accessibility of general and specialized obstetric care providers in Germany and England: an analysis of location and neonatal outcome. *International journal of health geographics,* 16**,** 44.

Benson, D. F. (1978) Neuropsychological Assessment By Muriel Lezak.(Pp. 549;£ 8.50.) Oxford University Press: London. 1976. *Psychological Medicine,* 8**,** 347-349.

Bernasko, J., L. Lynch, R. Lapinski & R. L. Berkowitz (1997) Twin pregnancies conceived by assisted reproductive techniques: maternal and neonatal outcomes. *Obstetrics & Gynecology,* 89**,** 368-372.

Bhattacharya, S. & D. M. Campbell (2005) The incidence of severe complications of preeclampsia. *Hypertension in pregnancy,* 24**,** 181-190.

Bhattacharya, S., J. Townend & S. Bhattacharya (2010) Recurrent miscarriage: Are three miscarriages one too many? Analysis of a Scottish population-based database of 151,021 pregnancies. *European Journal of Obstetrics & Gynecology and Reproductive Biology,* 150**,** 24-27.

Blackwell, S. C., C. Gyamfi-Bannerman, J. R. Biggio, Jr., S. P. Chauhan, B. L. Hughes, J. M. Louis, T. Manuck, H. S. Miller, A. F. Das, R. Birch & M. J. Jozwiakowski (2018) PROLONG Clinical Study Protocol: Hydroxyprogesterone Caproate to Reduce Recurrent Preterm Birth. *Am J Perinatol,* 35**,** 1228-1234.

Blencowe, H., S. Cousens, D. Chou, M. Oestergaard, L. Say, A.-B. Moller, M. Kinney & J. Lawn (2013) Born too soon: the global epidemiology of 15 million preterm births. *Reproductive health,* 10**,** S2.

Blencowe, H., S. Cousens, M. Z. Oestergaard, D. Chou, A.-B. Moller, R. Narwal, A. Adler, C. Vera Garcia, S. Rohde, L. Say & J. E. Lawn (2012) National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet,* 379**,** 2162-2172.

Blotkamp, A., D. Cromwell, B. Dumbrill, I. Gurol-Urganci, T. Harris, J. Hawdon & J. Jardine. 2018. National Maternity and Perinatal Audit: Clinical Report 2017-revised version. Healthcare Quality Improvement Partnership (HQIP).

Booth, A., M. Clarke, G. Dooley, D. Ghersi, D. Moher, M. Petticrew & L. Stewart (2012) The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Systematic reviews,* 1**,** 2.

Boulvain, M., C. M. Stan & O. Irion (2001) Elective delivery in diabetic pregnant women. *Cochrane database of systematic reviews*. 2001;(2):CD001997

Boyle, B., R. McConkey, E. Garne, M. Loane, M. Addor, M. Bakker, P. Boyd, M. Gatt, R. Greenlees, M. Haeusler, K. Klungsøyr, A. Latos-Bielenska, N. Lelong, R. McDonnell, J. Métneki, C. Mullaney, V. Nelen, M. O' Mahony, A. Pierini, J. Rankin, A. Rissmann, D. Tucker, D. Wellesley & H. Dolk (2013) Trends in the prevalence, risk and pregnancy outcome of multiple births with congenital anomaly: a registry-based study in 14 E uropean countries 1984–2007. *BJOG: An International Journal of Obstetrics & Gynaecology,* 120**,** 707-716.

Brady, A. R. (1998) Adjusted population attributable fractions from logistic regression. *Stata Technical Bulletin,* 7.

Butler, A. S. & R. E. Behrman. 2007. *Preterm birth: causes, consequences, and prevention*. National Academies Press.

Campbell, S. & P. Soothill (1993) Detection and management of intrauterine growth retardation: a British approach. *Ultrasound in obstetrics and gynaecology,* 2**,** 1432-5.

Carlin, J. B., L. C. Gurrin, J. A. Sterne, R. Morley & T. Dwyer (2005) Regression models for twin studies: a critical review. *International journal of epidemiology,* 34**,** 1089-1099.

Carroll, S. (2010) Preterm prelabour rupture of membranes (RCOG green-top guideline no. 44).

Chamberlain, G. (1991) Multiple pregnancy. *British Medical Journal,* 303**,** 111.

Chan, E., P. Leong, R. Malouf & M. Quigley (2016) Long-term cognitive and school outcomes of late-preterm and early-term births: a systematic review. *Child: care, health and development,* 42**,** 297-312.

Chang, H. H., J. Larson, H. Blencowe, C. Y. Spong, C. P. Howson, S. Cairns-Smith, E. M. Lackritz, S. K. Lee, E. Mason & A. C. Serazin (2013) Preventing preterm births: analysis of trends and potential reductions with interventions in 39 countries with very high human development index. *The Lancet,* 381**,** 223-234.

Cheong-See, F., E. Schuit, D. Arroyo-Manzano, A. Khalil, J. Barrett, K. Joseph, E. Asztalos, K. Hack, L. Lewi & A. Lim (2016) Prospective risk of stillbirth and neonatal complications in twin pregnancies: systematic review and meta-analysis. *British medical journal,* 354**,** i4353.

Cnattingius, S., A. Ericson, J. Gunnarskog & B. Källén (1990) A quality study of a medical birth registry. *Scandinavian journal of social medicine,* 18**,** 143-148.

Cole, S. (1980) Scottish maternity and neonatal records. *Perinatal audit and surveillance. London: Royal College of Obstetricians and Gynaecologists***,** 39-51.

Costeloe, K. L., E. M. Hennessy, S. Haider, F. Stacey, N. Marlow & E. S. Draper (2012) Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *British Medical Journal,* 345**,** e7976.

Crowther, C. A., P. Ashwood, A. J. McPhee, V. Flenady, T. Tran, J. M. Dodd, J. S. Robinson & P. S. Group (2017) Vaginal progesterone pessaries for pregnant women with a previous preterm birth to prevent neonatal respiratory distress syndrome (the PROGRESS Study): A multicentre, randomised, placebo-controlled trial. *PLoS medicine,* 14**,** e1002390.

Cui, J. (2007) QIC program and model selection in GEE analyses. *Stata journal,* 7**,** 209.

Davies, E. L., J. S. Bell & S. Bhattacharya (2016) Preeclampsia and preterm delivery: A population-based case–control study. *Hypertension in pregnancy,* 35**,** 510-519.

Davis, E. P., C. Buss, T. Muftuler, K. Head, A. Hasso, D. Wing, C. Hobel & C. A. Sandman (2011) Children's brain development benefits from longer gestation. *Frontiers in psychology,* 2**,** 1.

De Mauro, A., M. Greco & M. Grimaldi (2016) A formal definition of Big Data based on its essential features. *Library Review,* 65**,** 122-135.

Delnord, M., B. Blondel & J. Zeitlin (2015) What contributes to disparities in the preterm birth rate in European countries? *Current opinion in obstetrics & gynecology,* 27**,** 133.

De-Regil, L. M., C. Palacios, L. K. Lombardo & J. P. Peña-Rosas (2016) Vitamin D
supplementation for women during pregnancy. *Cochrane Database of Systematic Reviews*.
148-155. doi: 10.1016

Dhanjal, M. & A. Kenyon (2013) Induction of labour at term in older mothers. *RCOG Scientific Impact Paper No. 34*. RCOG, London

Di Renzo, G. C., I. Giardina, A. Rosati, G. Clerici, M. Torricelli, F. Petraglia & I. P. N. S. Group (2011) Maternal risk factors for preterm birth: a country-based population analysis. *European Journal of Obstetrics & Gynecology and Reproductive Biology,* 159**,** 342-346.

Dibben, C. & T. Clemens (2015) Place of work and residential exposure to ambient air pollution and birth outcomes in Scotland, using geographically fine pollution climate mapping estimates. *Environmental research,* 140**,** 535-541.

Dodd, J., C. Crowther, R. Haslam & J. Robinson (2012) Twins Timing of Birth Trial. Elective birth at 37 weeks of gestation versus standard care for women with an uncomplicated twin pregnancy at term: the Twins Timing of Birth Randomised Trial. *BJOG: An International Journal of Obstetrics & Gynaecology,* 119**,** 964-73.

Dodd, J. M., L. Jones, V. Flenady, R. Cincotta & C. A. Crowther (2013) Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. *Cochrane Database of Systematic Reviews*. CD004947

Doss, A. E., M. S. Mancuso, S. P. Cliver, V. C. Jauk & S. M. Jenkins (2012) Gestational age at delivery and perinatal outcomes of twin gestations. *American Journal of Obstetrics and Gynecology,* 207**,** 410.e1-410.e6.

Duckitt, K. & D. Harrington (2005) Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. *British Medical Journal,* 330**,** 565.

Eide, M. G., N. Øyen, R. Skjærven & T. Bjerkedal (2007) Associations of birth size, gestational age, and adult size with intellectual performance: evidence from a cohort of Norwegian men. *Pediatric research,* 62**,** 636.

Espel, E. V., L. M. Glynn, C. A. Sandman & E. P. Davis (2014) Longer gestation among children born full term influences cognitive and motor development. *PLoS One,* 9**,** e113758.

Ferrero, D. M., J. Larson, B. Jacobsson, G. C. Di Renzo, J. E. Norman, J. N. Martin, M. D'Alton, E. Castelazo, C. P. Howson, V. Sengpiel, M. Bottai, J. A. Mayo, G. M.

Shaw, I. Verdenik, N. Tul, P. Velebil, S. Cairns-Smith, H. Rushwan, S. Arulkumaran, J. L. Howse & J. L. Simpson (2016) Cross-Country Individual Participant Analysis of 4.1 Million Singleton Births in 5 Countries with Very High Human Development Index Confirms Known Associations but Provides No Biologic Explanation for 2/3 of All Preterm Births.(Research Article). *PLoS ONE,* 11**,** e0162506.

Fields, E., J. Chard, D. James & T. Treasure (2013) Fertility (update): summary of NICE guidance. *British Medical Journal,* 346**,** 33.

Figlio, D. N., J. Guryan, K. Karbownik & J. Roth (2016) Long-term cognitive and health outcomes of school-aged children who were born late-term vs full-term. *JAMA pediatrics,* 170**,** 758-764.

Fleming, M., C. A. Fitton, M. F. Steiner, J. S. McLay, D. Clark, A. King, D. F. Mackay & J. P. Pell (2017) Educational and health outcomes of children treated for attention-deficit/hyperactivity disorder. *JAMA pediatrics,* 171**,** e170691-e170691.

Flenady, V., G. Hawley, O. M. Stock, S. Kenyon & N. Badawi (1998) Prophylactic antibiotics for inhibiting preterm labour with intact membranes. *Cochrane Pregnancy and Childbirth Group,* 2013.CD 000246

Fuchs, F. & M. V. Senat (2016) Multiple gestations and preterm birth. *Seminars in Fetal and Neonatal Medicine,* 21**,** 113-120.

Gates, S. & P. Brocklehurst (2004) How should randomised trials including multiple pregnancies be analysed? *BJOG: An International Journal of Obstetrics & Gynaecology,* 111**,** 213-219.

Gibson, K. S., T. P. Waters, D. D. Gunzler & P. M. Catalano (2015) A retrospective cohort study of factors relating to the longitudinal change in birth weight. *BMC pregnancy and childbirth,* 15**,** 344.

Goldenberg, R. L., J. F. Culhane, J. D. Iams & R. Romero (2008) Epidemiology and causes of preterm birth. *The lancet,* 371**,** 75-84.

Goya, M., M. de La Calle, L. Pratcorona, C. Merced, C. Rodó, B. Muñoz, M. Juan, A. Serrano, E. Llurba, T. Higueras, E. Carreras & L. Cabero (2016) Cervical pessary to prevent preterm birth in women with twin gestation and sonographic short cervix: a multicenter randomized controlled trial (PECEP-Twins). *American Journal of Obstetrics and Gynecology,* 214**,** 145-152.

Goya, M., L. Pratcorona, C. Merced, C. Rodó, L. Valle, A. Romero, M. Juan, A. Rodríguez, B. Muñoz, B. Santacruz, J. C. Bello-Muñoz, E. Llurba, T. Higueras, L. Cabero & E. Carreras (2012) Cervical pessary in pregnant women with a short cervix (PECEP): an open- label randomised controlled trial. *The Lancet,* 379**,** 1800-1806.

Greene, M. F. (2018). Choices in managing full-term pregnancy. *New England Journal of Medicine* 379: 580-581.

Grobman, W. (2019) Resource utilization among low-risk nulliparas randomized to elective induction at 39 weeks or expectant management. *American Journal of Obstetrics and Gynecology,* 220 S(S2-S3)

Grobman, W. A. & A. B. Caughey (2019) Elective induction of labor at 39 weeks compared to expectant management: A meta-analysis of cohort studies. *American journal of obstetrics and gynecology*. DOI:/10.1016/ajog.2019.

Grobman, W. A., M. M. Rice, U. M. Reddy, A. T. Tita, R. M. Silver, G. Mallett, K. Hill, E. A. Thom, Y. Y. El-Sayed & A. Perez-Delboy (2018) Labor induction versus expectant management in low-risk nulliparous women. *New England Journal of Medicine,* 379**,** 513-523.

Grove, B. J., S. J. Lim, C. R. Gale & S. D. Shenkin (2017) Birth weight and cognitive ability in adulthood: A systematic review and meta-analysis. *Intelligence,* 61**,** 146-158.

Grömping, U. (2006) Relative importance for linear regression in R: the package relaimpo. *Journal of statistical software,* 17**,** 1-27.

Gyamfi-Bannerman, C. 2011. The scope of the problem: the epidemiology of late preterm and early-term birth. In *Seminars in perinatology*, 246-248. Elsevier.
Gyamfi-Bannerman, C., M. Son & C. Ananth (2014) 696: Long-term neurodevelopmental outcomes of late preterm and early term neonates. *American Journal of Obstetrics & Gynecology,* 210**,** S342-S343.
Gyetvai, E. K., D. M. Hannah, D. E. Hodnett & D. A. Ohlsson (1999) Tocolytics for preterm labor: A systematic Review. *Obstetrics & Gynecology,* 94**,** 869-877.
Gülmezoglu, A. M., C. A. Crowther, P. Middleton & E. Heatley (2012) Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane database of systematic reviews*. CD004945
Haas, D. M., D. M. Caldwell, P. Kirkpatric, J. J. McIntosh & N. J. Welton (2012) Tocolytic therapy for preterm delivery: systematic review and network meta-analysis.(Clinical report). *British Medical Journal,* 345**,** 17.
Hanley, J. A., A. Negassa, M. D. d. Edwardes & J. E. Forrester (2003) Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation. *American Journal of Epidemiology,* 157**,** 364-375.
Hannah, E. K., A. C. David, G.-U. Ipek, H. Katie, H. v. D. M. Jan & C. S. S. Gordon (2017) Perinatal mortality associated with induction of labour versus expectant management in nulliparous women aged 35 years or over: An English national cohort study. *PLoS Medicine,* 14**,** e1002425.
Hao, H., H. H. Chang, H. A. Holmes, J. A. Mulholland, M. Klein, L. A. Darrow & M. J. Strickland (2015) Air pollution and preterm birth in the US State of Georgia (2002–2006): associations with concentrations of 11 ambient air pollutants estimated by combining Community Multiscale Air Quality Model (CMAQ) simulations with stationary monitor measurements. *Environmental health perspectives,* 124**,** 875-880.
Harrison, W. & D. Goodman (2015) Epidemiologic trends in neonatal intensive care, 2007-2012. *JAMA pediatrics,* 169**,** 855-862.
Heazell, A. E. P., M. K. Whitworth, J. Whitcombe, S. W. Glover, C. Bevan, J. Brewin, C. Calderwood, A. Canter, F. Jessop, G. Johnson, I. Martin & L. Metcalf (2015) Research priorities for stillbirth: process overview and results from UK Stillbirth Priority Setting Partnership. *Ultrasound in Obstetrics & Gynecology,* 46**,** 641-647.
Helmerhorst, F. M., D. A. Perquin, D. Donker & M. J. Keirse (2004) Perinatal outcome of singletons and twins after assisted conception: a systematic review of controlled studies. *British Medical Journal,* 328**,** 261.
Henderson, J., C. Hockley, S. Petrou, M. Goldacre & L. Davidson (2004) Economic implications of multiple births: inpatient hospital costs in the first 5 years of life. *Archives of Disease in Childhood-Fetal and Neonatal Edition,* 89**,** F542-F545.
Hilder, L., K. Costeloe & B. Thilaganathan (1998) Prolonged pregnancy: evaluating gestation-specific risks of fetal and infant mortality. *BJOG: An International Journal of Obstetrics & Gynaecology,* 105**,** 169-173.
Hosmer Jr, D. W. & S. Lemeshow (1999) Applied survival analysis: regression modelling of time to event data (1999). *Eur Orthodontic Soc***,** 561-2.
Humphrey, T. & J. S. Tucker (2009) Rising rates of obstetric interventions: exploring the determinants of induction of labour. *Journal of Public Health,* 31**,** 88-94.
Huttenlocher, P. R. & A. S. Dabholkar (1997) Regional differences in synaptogenesis in human cerebral cortex. *Journal of comparative Neurology,* 387**,** 167-178.
Ibanez, G., M.-A. Charles, A. Forhan, G. Magnin, O. Thiebaugeorges, M. Kaminski, M.-J. Saurel-Cubizolles & E. M. C. C. S. Group (2012) Depression and anxiety in women during pregnancy and neonatal outcome: data from the EDEN mother–child cohort. *Early human development,* 88**,** 643-649.
ISD Scotland Statistics. Births in Scottish Maternity Hospital, 2018. Scotland, Edinburgh.

ISD Scotland, 2009. Trends in Perinatal Mortality in Scotland: a review over 30 years. NHS Quality Improvement Scotland, Edinburgh.

Jarde, A., O. Lutsiv, C. Park, J. Beyene, J. Dodd, J. Barrett, P. Shah, J. Cook, S. Saito & A. Biringer (2017) Effectiveness of progesterone, cerclage and pessary for preventing preterm birth in singleton pregnancies: a systematic review and network meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology,* 124**,** 1176-1189.

Joy, J., N. McClure & I. Cooke (2008) A comparison of spontaneously conceived twins and twins conceived by artificial reproductive technologies. *Journal of Obstetrics and Gynaecology,* 28**,** 580-585.

Kahn, B., L. Lumey, P. A. Zybert, J. M. Lorenz, J. Cleary-Goldman, M. E. D'alton & J. N. Robinson (2003) Prospective risk of fetal death in singleton, twin, and triplet gestations: implications for practice. *Obstetrics & Gynecology,* 102**,** 685-692.

Kenyon, S., M. Boulvain & J. P. Neilson (2010) Antibiotics for preterm rupture of membranes. *Cochrane database of systematic reviews (Online),* 8**,** CD001058.

Kenyon, S., H. Hagberg & J. Norman (2013) Preterm labour, antibiotics, and cerebral palsy. *United Kingdom: RCOG Scientific Impact Paper No. 33.*RCOG, London

Kenyon, S., K. Pike, D. Jones, P. Brocklehurst, N. Marlow, A. Salt & D. Taylor (2008) Childhood outcomes after prescription of antibiotics to pregnant women with preterm rupture of the membranes: 7-year follow-up of the ORACLE I trial. *The Lancet,* 372**,** 1310-1318.

Kenyon, S., D. Taylor & W. Tarnow-Mordi (2001) Broad-spectrum antibiotics for preterm, prelabour rupture of fetal membranes: the ORACLE I randomised trial. *The Lancet,* 357**,** 979-988.

Kernell, K., G. Sydsjö, M. Bladh, N. E. Nielsen & A. Josefsson (2014) Congenital heart disease in men–birth characteristics and reproduction: a national cohort study. *BMC pregnancy and childbirth,* 14**,** 187.

Kilby, M. D. (2017) Management of Monochorionic Twin Pregnancy: Green-top Guideline No. 51. *BJOG : an international journal of obstetrics and gynaecology,* 124**,** e1.

Kim, S. Y., J. E. Park, Y. J. Lee, H.-J. Seo, S.-S. Sheen, S. Hahn, B.-H. Jang & H.-J. Son (2013) Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *Journal of clinical epidemiology,* 66**,** 408-414.

Kirkwood, B. R. & J. A. Sterne. 2010. *Essential medical statistics*. John Wiley & Sons.

Kramer, M. S., S. Liu, Z. Luo, H. Yuan, R. W. Platt & K. Joseph (2002) Analysis of perinatal mortality and its components: time for a change? *American Journal of Epidemiology,* 156**,** 493-497.

Liem, S., E. Schuit, M. Hegeman, J. Bais, K. de Boer, K. Bloemenkamp, J. Brons, H. Duvekot, B. N. Bijvank, M. Franssen, I. Gaugler, I. de Graaf, M. Oudijk, D. Papatsonis, P. Pernet, M. Porath, L. Scheepers, M. Sikkema, J. Sporken, H. Visser, W. van Wijngaarden, M. Woiski, M. van Pampus, B. W. Mol & D. Bekedam (2013a) Cervical pessaries for prevention of preterm birth in women with a multiple pregnancy (ProTWIN): a multicentre, open-label randomised controlled trial. *The Lancet,* 382.

Liem, S. M. S., M. G. van Pampus, B. W. J. Mol & D. J. Bekedam (2013b) Cervical Pessaries for the Prevention of Preterm Birth: A Systematic Review. *Obstetrics and Gynecology International,* 2013.

Link, G., K. E. Clark & U. Lang (2007) Umbilical blood flow during pregnancy: evidence for decreasing placental perfusion. *American journal of obstetrics and gynecology,* 196**,** 489. e1-489. e7.

Liu, S., D. Krewski, Y. Shi, Y. Chen & R. T. Burnett (2003) Association between gaseous ambient air pollutants and adverse pregnancy outcomes in Vancouver, Canada. *Environmental health perspectives,* 111**,** 1773-1778.

Lunt, M. (2011) A guide to imputing missing data with Stata: Revision 1.4. http://personalpages.manchester.ac.uk/staff/mark.lunt/mi_guide.pdf

Ma, J. P., S. Huang, S. Qin, C. You & Y. Zeng (2016) Progesterone for acute traumatic brain injury. *Cochrane Database Syst Rev.,* 12. CD 008489.

MacKay, D. F., G. C. S. Smith, R. Dobbie & J. P. Pell (2010) Gestational Age at Delivery and Special Educational Need: Retrospective Cohort Study of 407,503 Schoolchildren (Gestation and Special Educational Need). *PLoS Medicine,* 7**,** e1000289.

Manktelow, B., L. Smith, S. Seaton, P. Hyman-Taylor, J. Kurinczuk, D. Field, P. Smith & E. Draper (2014) Collaboration. obotM-U: MBRRACE-UK Perinatal mortality surveillance report. *UK Perinatal deaths for births from January to December 2014*.

Maria, K., K. Paraskevi, I. Zoi, C. Ilias, P. Eleni, E. Sotiria & N. Nikolaos (2013) Comparison of 24 months neurodevelopmental outcome in twins and singletons ≤ 34 weeks gestation at birth. *Journal of Pediatric and Neonatal Individualized Medicine,* 2**,** 48-54.

Martínez-Nadal, S., X. Demestre, F. Raspall, J. Alvarez, M. Elizari, C. Vila & P. Sala (2014) Neonatal morbidity in early-term newborns. *Anales de Pediatría (English Edition),* 81**,** 39-44.

McDonald, S., K. Murphy, J. Beyene & A. Ohlsson (2005) Perinatal outcomes of in vitro fertilization twins: a systematic review and meta-analyses. *American journal of obstetrics and gynecology,* 193**,** 141-152.

McNamara, H. C., R. Wood, J. Chalmers, N. Marlow, J. Norrie, G. MacLennan, G. McPherson, C. Boachie & J. E. Norman (2015) STOPPIT Baby Follow-Up Study:The Effect of Prophylactic Progesterone in Twin Pregnancy on Childhood Outcome. *PlosOne. DOI: 10.1371/journal.pone.0122341*.

Meis, P. J., M. Klebanoff, E. Thom, M. P. Dombrowski, B. Sibai, A. H. Moawad, C. Y. Spong, J. C. Hauth, M. Miodovnik & M. W. Varner (2003) Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *New England Journal of Medicine,* 348**,** 2379-2385.

Merced, C.,Goya, L. Pratcorona, C. Rodó, E. Llurba, T. Higueras, L. Cabero, E. Carreras & T. Group (2019) Cervical pessary for preventing preterm birth in twin pregnancies with maternal short cervix after an episode of threatened preterm labour: randomised controlled trial. *American journal of obstetrics and gynecology*. 2016;214(2):145-152**.**

Min, S.-J., B. Luke, B. Gillespie, L. Min, R. B. Newman, J. G. Mauldin, F. R. Witter, F. A. Salman & M. J. O'sullivan (2000) Birth weight references for twins. *American journal of obstetrics and gynecology,* 182**,** 1250-1257.

Minakami, H. & I. Sato (1996) Reestimating Date of Delivery in Multifetal Pregnancies. *JAMA,* 275**,** 1432-1434.

Moher, D., A. Liberati, J. Tetzlaff & D. G. Altman (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine,* 151**,** 264-269.

Moore, T., E. M. Hennessy, J. Myles, S. J. Johnson, E. S. Draper, K. L. Costeloe & N. Marlow (2012) Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. *British Medical Journal,* 345**,** e7961.

Morken, N.-H., K. Källen, H. Hagberg & B. Jacobsson (2005) Preterm birth in Sweden 1973-2001: rate, subgroups, and effect of changing patterns in multiple births, maternal age, and smoking. *Acta obstetricia et gynecologica Scandinavica,* 84**,** 558-565.

Morken, N.-H., K. Källen & B. Jacobsson (2006) Fetal growth and onset of delivery: a nationwide population-based study of preterm infants. *American journal of obstetrics and gynecology,* 195**,** 154-161.

Morris, R. & V. Carstairs (1991) Which deprivation? A comparison of selected deprivation indexes. *Journal of Public Health,* 13**,** 318-326.

Murray, S. R., S. J. Stock & J. E. Norman. 2017. Long-term childhood outcomes after interventions for prevention and management of preterm birth. In *Seminars in perinatology*, 519-527.

Naimi, A. & N. Auger (2016) Cumulative risk of stillbirth in the presence of competing events. *BJOG: An International Journal of Obstetrics & Gynaecology,* 123**,** 1071-1074.

Napierala, M. A. (2012) What is the Bonferroni correction. *AAOS Now,* 6**,** 40.

Narberhaus, A., D. Segarra, X. Caldú, M. Giménez, C. Junqué, R. Pueyo & F. Botet (2007) Gestational age at preterm birth in relation to corpus callosum and general cognitive outcome in adolescents. *Journal of child neurology,* 22**,** 761-765.

Nepomnyaschy, L., T. Hegyi, B. M. Ostfeld & N. E. Reichman (2012) Developmental outcomes of late-preterm infants at 2 and 4 years. *Maternal and child health journal,* 16**,** 1612-1624.

National Institute of Health and Care Excellence 2008. *Antenatal care: routine care for the healthy pregnant woman*. NICE: London

National Institute for Health and Care Excellence, (2010) Antenatal Care. NICE Clinical Guideline 62. NICE: London.

Network RCOG. A report by the National Guideline Alliance about twin pregnancy costing. 2017 RCOG: London

Newcombe, H. B. & J. M. Kennedy (1962) Record linkage: making maximum use of the discriminating power of identifying information. *Communications of the ACM,* 5**,** 563-566.

Nicolaides, K. H., A. Syngelaki, L. C. Poon, C. de Paco Matallana, W. Plasencia, F. S. Molina, G. Picciarelli, N. Tul, E. Celik, T. K. Lau & R. Conturso (2016) Cervical pessary placement for prevention of preterm birth in unselected twin pregnancies: a randomized controlled trial. *American Journal of Obstetrics and Gynecology,* 214**,** 3.e1-3.e9.

Norman, J. E. & P. Bennett (2017) Preterm birth prevention—Time to PROGRESS beyond progesterone. *PLoS medicine,* 14**,** e1002391.

Norman, J. E., A. E. Heazell, A. Rodriguez, C. J. Weir, S. J. Stock, C. J. Calderwood, S. C. Burley, J. F. Frøen, M. Geary & F. Breathnach (2018) Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM): a stepped wedge, cluster-randomised trial. *The Lancet,* 392**,** 1629-1638.

Norman, J. E., F. Mackenzie, P. Owen, H. Mactier, K. Hanretty, S. Cooper, A. Calder, G. Mires, P. Danielian, S. Sturgiss, G. Maclennan, G. Tydeman, S. Thornton, B. Martin, J. G. Thornton, J. P. Neilson & J. Norrie (2009) Progesterone for the prevention of preterm birth in twin pregnancy (STOPPIT): a randomised, double-blind, placebo-controlled study and meta-analysis. *The Lancet,* 373**,** 2034-2040.

Norman, J. E., N. Marlow, C.-M. Messow, A. Shennan, P. R. Bennett, S. Thornton, S. C. Robson, A. McConnachie, S. Petrou & N. J. Sebire (2016) Vaginal progesterone prophylaxis for preterm birth (the OPPTIMUM study): a multicentre, randomised, double-blind trial. *The lancet,* 387**,** 2106-2116.

Northen, A. T., G. S. Norman, K. Anderson, L. Moseley, M. DiVito, M. Cotroneo, M. Swain, S. Bousleiman, F. Johnson & K. Dorman (2007) Follow-up of children exposed in utero to 17 α-hydroxyprogesterone caproate compared with placebo. *Obstetrics & Gynecology,* 110**,** 865-872.

Norwitz, E. R., J. N. Robinson & J. R. Challis (1999) The control of labor. *New England Journal of Medicine,* 341**,** 660-666.

O'brien, J. (2015) Progestogen safety in multiple gestations: application of the Bradford Hill criteria. *BJOG: An International Journal of Obstetrics & Gynaecology,* 122**,** 610-614.

O'brien, J. M. (2012) The safety of progesterone and 17-hydroxyprogesterone caproate administration for the prevention of preterm birth: an evidence-based assessment. *American journal of perinatology,* 29**,** 665-672.

Royal College of Obstetricians and Gynaecologists. 2015. Cervical cerclage. Green–top Guideline No. 60. RCOG: London.

Olesen, A. W., O. Basso & J. Olsen (2003) Risk of recurrence of prolonged pregnancy. *British Medical Journal,* 326**,** 476.

O'Brien, J. M. & D. F. Lewis (2016) Prevention of preterm birth with vaginal progesterone or 17-alpha-hydroxyprogesterone caproate: a critical examination of efficacy and safety. *American journal of obstetrics and gynecology,* 214**,** 45-56.

Pandey, S., A. Shetty, M. Hamilton, S. Bhattacharya & A. Maheshwari (2012) Obstetric and perinatal outcomes in singleton pregnancies resulting from IVF/ICSI: a systematic review and meta-analysis. *Human reproduction update,* 18**,** 485-503.

Peter, C., P. Wenzlaff, J. Kruempelmann, G. Alzen, E. Bueltmann & S. E. Gruessner (2013) Perinatal morbidity and early neonatal mortality in twin pregnancies. *Open Journal of Obstetrics and Gynecology,* 3**,** 78.

Pintilie, M. 2006. *Competing risks: a practical perspective*. John Wiley & Sons.

Popay, J., H. Roberts, A. Sowden, M. Petticrew, L. Arai, M. Rodgers, N. Britten, K. Roen & S. Duffy (2006) Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version,* 1**,** b92.

Pérez-López, F. R., P. Chedraui, G. R. Pérez-Roncero & S. J. Martínez-Domínguez (2019) Effectiveness of the cervical pessary for the prevention of preterm birth in singleton pregnancies with a short cervix: a meta-analysis of randomized trials. *Archives of gynecology and obstetrics***,** 1-17.

Rafael, T., V. Berghella & Z. Alfirevic (2014) Cervical stitch (cerclage) for preventing preterm birth in multiple pregnancy. *Cochrane Database Syst Rev.,* 2014. CD009196.

Reddy, U. M., C.-W. Ko, T. N. Raju & M. Willinger (2009) Delivery indications at late-preterm gestations and infant mortality rates in the United States. *Pediatrics,* 124**,** 234.

Reddy, U. M., C.-W. Ko & M. Willinger (2006) Maternal age and the risk of stillbirth throughout pregnancy in the United States. *American journal of obstetrics and gynecology,* 195**,** 764-770.

Ressel, G. W. (2004) ACOG releases bulletin on managing cervical insufficiency.(Practice Guidelines)(American College of Obstetricians and Gynecologists). *American Family Physician,* 69**,** 436.

Reynolds, R. M., K. M. Allan, E. A. Raja, S. Bhattacharya, G. McNeill, P. C. Hannaford, N. Sarwar, A. J. Lee, S. Bhattacharya & J. E. Norman (2013) Maternal obesity during pregnancy and premature mortality from cardiovascular event in adult offspring: follow-up of 1 323 275 person years. *British Medical Journal,* 347**,** f4539.

Rode, L., K. Klein, K. Nicolaides, E. Krampl-Bettelheim, A. Tabor & P. Group (2011) Prevention of preterm delivery in twin gestations (PREDICT): a multicenter, randomized, placebo-controlled trial on the effect of vaginal micronized progesterone. *Ultrasound in Obstetrics & Gynecology,* 38**,** 272-280.

Romeo, D. M., C. Brogna, F. Sini, M. G. Romeo, F. Cota & D. Ricci (2016) Early psychomotor development of low-risk preterm infants: Influence of gestational age and gender. *European Journal of Paediatric Neurology,* 20**,** 518-523.

Romero, R., S. K. Dey & S. J. Fisher (2014a) Preterm labor: one syndrome, many causes. *Science,* 345**,** 760-765.

Romero, R., S. S. Hassan, P. Gajer, A. L. Tarca, D. W. Fadrosh, L. Nikita, M. Galuppi, R. F. Lamont, P. Chaemsaithong, J. Miranda, T. Chaiworapongsa & J. Ravel (2014b) The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome,* 2**,** 4.

Romero, R., K. Nicolaides, A. Conde-Agudelo, A. Tabor, J. M. O'brien, E. Cetingoz, E. Da Fonseca, G. W. Creasy, K. Klein & L. Rode (2012) Vaginal progesterone in women with an asymptomatic sonographic short cervix in the midtrimester decreases preterm delivery and neonatal morbidity: a systematic review and metaanalysis of individual patient data. *American journal of obstetrics and gynecology,* 206**,** 124. e1-124. e19.

Romero, R., B. M. Sibai, L. Sanchez-Ramos, G. J. Valenzuela, J.-C. Veille, B. Tabor, K. G. Perry, M. Varner, T. M. Goodwin, R. Lane, J. Smith, G. Shangold & G. W. Creasy (2000) An oxytocin receptor antagonist (atosiban) in the treatment of preterm labor: A randomized, double-blind, placebo-controlled trial with tocolytic rescue. *American Journal of Obstetrics and Gynecology,* 182**,** 1173-1183.

Roos, C., M. E. A. Spaanderman, E. Schuit, K. W. M. Bloemenkamp, A. C. Bolte, J. Cornette, J. J. J. Duvekot, J. van Eyck, M. T. M. Franssen, C. J. de Groot, J. H. Kok, A. Kwee, A. Merién, B. N. Bijvank, B. C. Opmeer, M. A. Oudijk, M. G. van Pampus, D. N. M. Papatsonis, M. M. Porath, H. C. J. Scheepers, S. A. Scherjon, K. M. Sollie, S. M. C. Vijgen, C. Willekes, B. W. J. Mol, J. A. M. van Der Post & F. K. Lotgering (2013) Effect of Maintenance Tocolysis With Nifedipine in Threatened Preterm Labor on Perinatal Outcomes: A Randomized Controlled Trial. *JAMA,* 309**,** 41-47.

Rose, O., E. Blanco, S. M. Martinez, E. K. Sim, M. Castillo, B. Lozoff, Y. E. Vaucher & S. Gahagan (2013) Developmental scores at 1 year with increasing gestational age, 37–41 weeks. *Pediatrics,* 131**,** e1475.

Royal College of Obstetricians and Gynaecologists. (2010) Greentop guideline No. 7: Antenatal corticosteroids to reduce neonatal morbidity and mortality. London: RCOG.

Royston, P. (2007) Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. *Stata Journal,* 7**,** 445-464.

Rubin, D. B. 2004. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.

Räisänen, S., M. Gissler, J. Saari, M. Kramer & S. Heinonen (2013) Contribution of risk factors to extremely, very and moderately preterm births–register-based analysis of 1,390,742 singleton births. *PLoS One,* 8**,** e60660.

Saccone, G., G. M. Maruotti, A. Giudicepietro & P. Martinelli (2017) Effect of Cervical Pessary on Spontaneous Preterm Birth in Women With Singleton Pregnancies and Short Cervical Length: A Randomized Clinical Trial. *JAMA,* 318**,** 2317-2324.

Saccone, G., O. Rust, S. Althuisius, A. Roman & V. Berghella. 2015. Cerclage for short cervix in twin pregnancies: systematic review and meta-analysis of randomized trials using individual patient-level data. *Acta Obstet Gynecol Scand*. 2015;94(4)352-358.

Sairam, S., K. Costeloe & B. Thilaganathan (2002) Prospective risk of stillbirth in multiple-gestation pregnancies: a population-based analysis. *Obstetrics & Gynecology,* 100**,** 638-641.

Sarri, G., M. Davies, M. Gholitabar & J. E. Norman (2015) Preterm labour: summary of NICE guidance. *British Medical Journal,* 351.

Schuit, E., S. Stock, L. Rode, D. Rouse, A. Lim, J. Norman, A. Nassar, V. Serra, C. Combs, C. Vayssiere, M. Aboulghar, S. Wood, E. Çetingöz, C. Briery, E. Fonseca, K. Worda, A. Tabor, E. Thom, S. Caritis, J. Awwad, I. Usta, A. Perales, J. Meseguer, K. Maurel, T. Garite, M. Aboulghar, Y. Amin, S. Ross, C. Cam, A. Karateke, J. Morrison, E. Magann, K. Nicolaides, N. Zuithoff, R. Groenwold, K. Moons, A. Kwee & B. Mol.

2015. Effectiveness of progestogens to improve perinatal outcome in twin pregnancies: an individual participant data meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*: 122: 27-37.

Singh, M. & C. Su (2013) Progesterone and neuroprotection. *Hormones and behavior,* 63**,** 284.

Smith, G. C. (2001) Life-table analysis of the risk of perinatal death at term and post term in singleton pregnancies. *American journal of obstetrics and gynecology,* 184**,** 489-496.

Smith G.C.(2005) Estimating risks of perinatal death. *American journal of obstetrics and gynecology,* 192**,** 17-22.

Smith, G. C., K. M. Fleming & I. R. White (2007) Birth order of twins and risk of perinatal death related to delivery in England, Northern Ireland, and Wales, 1994-2003: retrospective cohort study. *British Medical Journal,* 334**,** 576.

Smith, G. C., J. P. Pell & R. Dobbie (2002) Birth order, gestational age, and risk of delivery related perinatal death in twins: retrospective cohort study. *British Medical Journal,* 325**,** 1004.

Smithers, L., A. Searle, C. Chittleborough, W. Scheil, S. Brinkman & J. Lynch (2015) A whole-of-population study of term and post-term gestational age at birth and children's development. *BJOG: An International Journal of Obstetrics & Gynaecology,* 122**,** 1303-1311.

Society of Maternal and Fetal Medicine (2014) ACOG Practice Bulletin No. 144: Multifetal gestations: twin, triplet, and higher-order multifetal pregnancies. *Obstetrics and gynecology,* 123**,** 1118.

Spong, C. Y., B. M. Mercer, M. D'Alton, S. Kilpatrick, S. Blackwell & G. Saade (2011) Timing of indicated late-preterm and early-term birth. *Obstetrics and gynecology,* 118**,** 323.

Stacey, T., J. M. Thompson, E. A. Mitchell, A. Ekeroma, J. Zuccollo & L. M. McCowan (2011) Maternal perception of fetal activity and late stillbirth risk: findings from the Auckland Stillbirth Study. *Birth,* 38**,** 311-316.

Stephens, B. E. & B. R. Vohr (2009) Neurodevelopmental outcome of the premature infant. *Pediatric Clinics,* 56**,** 631-646.

Sterne, J. A., I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood & J. R. Carpenter (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal,* 338**,** b2393.

Stewart, L. A., M. Simmonds, L. Duley, K. C. Dietz, M. Harden, A. Hodkinson, A. Llewellyn, S. Sharif, R. Walker & K. Wright (2017) Evaluating progestogens for prevention of preterm birth international collaborative (EPPPIC) individual participant data (IPD) meta-analysis: protocol. *Systematic reviews,* 6**,** 235.

Stock, S., E. Ferguson, A. Duffy, I. Ford, J. Chalmers & J. Norman (2012) Outcomes of elective induction of labour compared with expectant management: population based study. *British Medical Journal,* 344**,** 19.

Stock, S. & J. Norman. 2010. Preterm and term labour in multiple pregnancies. In *Seminars in Fetal and Neonatal Medicine*, 336-341. Elsevier.

Stock, S. J. & K. M. Ismail (2016) Which intervention reduces the risk of preterm birth in women with risk factors? *British Medical Journal ,* 355**,** i5206.

Suhag, A., G. Saccone, M. Bisulli, N. Seligman & V. Berghella (2015) Trends in cerclage use. *Acta Obstetricia et Gynecologica Scandinavica,* 94**,** 1188-1194.

Suzuki, S. & H. Miyake (2010) Perinatal outcomes of elderly primiparous dichorionic twin pregnancies conceived by in vitro fertilization compared with those conceived spontaneously. *Archives of gynecology and obstetrics,* 281**,** 87.

Suzuki, S., Y. Otsubo, R. Sawa, Y. Yoneyama & T. Araki (2000) Clinical trial of induction of labor versus expectant management in twin pregnancy. *Gynecologic and obstetric investigation,* 49**,** 24-27.

The International Classification of Diseases, 10th Edition (ICD-10) changeover is coming. *Optometry (St. Louis, Mo.),* 81**,** 551-553.

Teune, M. J., S. Bakhuizen, C. Gyamfi Bannerman, B. C. Opmeer, A. H. van Kaam, A. G. van Wassenaer, J. M. Morris & B. W. J. Mol (2011) A systematic review of severe morbidity in infants born late preterm. *American Journal of Obstetrics and Gynecology,* 205**,** 374.e1-374.e9.

Thérien, J. P. (2012) The United Nations and Human Development: From Ideology to Global Policies. *Global Policy,* 3**,** 1-12.

Tsou, M.-T., M.-W. Tsou, M.-P. Wu & J.-T. Liu (2008) Academic achievement of twins and singletons in early adulthood: Taiwanese cohort study. *British Medical Journal:,* 337.

van 't Hooft, J., J. H. van der Lee, B. C. Opmeer, A. G. van Wassenaer-Leemhuis, A. L. van Baar, D. J. Bekedam, L. J. P. Steenis, S. Liem, E. Schuit, C. Cuijpers, E. Bleeker, M. E. Vinke, N. Simons, I. M. de Graaf, B. W. J. Mol & C. van de Beek (2018) Pessary for prevention of preterm birth in twin pregnancy with short cervix: 3-year follow-up study.(Report). *Ultrasound in Obstetrics and Gynaecology,* 51**,** 621.

Van Buuren, S., H. C. Boshuizen & D. L. Knook (1999) Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in medicine,* 18**,** 681-694.

van den Broek, N. R., S. A. White, M. Goodall, C. Ntonya, E. Kayira, G. Kafulafula & J. P. Neilson (2009) The APPLe Study: A Randomized, Community-Based, Placebo-Controlled Trial of Azithromycin for the Prevention of Preterm Birth, with Meta-Analysis (Azithromycin to Prevent Preterm Birth). *PLoS Medicine,* 6**,** e1000191.

van Vliet, E. O. G., T. A. J. Nijman, E. Schuit, K. Y. Heida, B. C. Opmeer, M. Kok, W. Gyselaers, M. M. Porath, M. Woiski, C. J. Bax, K. W. M. Bloemenkamp, H. C. J. Scheepers, Y. Jacquemyn, E. v. Beek, J. J. Duvekot, M. T. M. Franssen, D. N. Papatsonis, J. H. Kok, J. A. M. van Der Post, A. Franx, B. W. Mol & M. A. Oudijk (2016) Nifedipine versus atosiban for threatened preterm birth (APOSTEL III): a multicentre, randomised controlled trial. *The Lancet,* 387**,** 2117-2124.

Vasak, B., J. J. Verhagen, S. V. Koenen, M. P. Koster, P. A. de Reu, A. Franx, J. G. Nijhuis, G. J. Bonsel & G. H. Visser (2017) Lower perinatal mortality in preterm born twins than in singletons: a nationwide study from The Netherlands. *American journal of obstetrics and gynecology,* 216**,** 161. e1-161. e9.

Vedel, C., H. Larsen, A. Holmskov, K. R. Andreasen, N. Uldbjerg, J. Ramb, B. Bødker, L. Skibsted, L. Sperling, L. Krebs, H. Zingenberg, L. Laursen, J. T. Christensen, A. Tabor & L. Rode (2016) Long-term effects of prenatal progesterone exposure: neurophysiological development and hospital admissions in twins up to 8 years of age. *Ultrasound in Obstetrics & Gynecology,* 48**,** 382-389.

Visintin, C., M. Mugglestone, D. James & M. Kilby (2011) Antenatal care for twin and triplet pregnancies: summary of NICE guidance. *British Medical Journal,* 343**,** 688.

Visintin, C., M. A. Mugglestone, M. Q. Almerie, L. M. Nherera, D. James & S. Walkinshaw (2010) Management of hypertensive disorders during pregnancy: summary of NICE guidance. *British Medical Journal,* 341**,** c2207.

Vliet, E., L. Seinen, C. Roos, E. Schuit, H. Scheepers, K. Bloemenkamp, J. Duvekot, J. Eyck, J. Kok, F. Lotgering, A. Baar, A. Wassenaer-Leemhuis, M. Franssen, M. Porath, J. Post, A. Franx, B. Mol & M. Oudijk (2016) Maintenance tocolysis with nifedipine in threatened preterm labour: 2-year follow up of the offspring in the APOSTEL II trial. *BJOG: An International Journal of Obstetrics & Gynaecology,* 123**,** 1107-1114.

Vohr, B. (2013) Long-term outcomes of moderately preterm, late preterm, and early term infants. *Clinics in perinatology,* 40**,** 739-751.

Von Elm, E., D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche & J. P. Vandenbroucke (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Annals of internal medicine,* 147**,** 573-577.

Walker, K. F., G. J. Bugg, M. Macpherson, C. McCormick, N. Grace, C. Wildsmith, L. Bradshaw, G. C. Smith & J. G. Thornton (2016) Randomized trial of labor induction in women 35 years of age or older. *New England Journal of Medicine,* 374**,** 813-822.

Whitworth, M., L. Bricker & C. Mullan (2015) Ultrasound for fetal assessment in early pregnancy. *Cochrane database of systematic reviews*. CD007058

Willing, J. & C. K. Wagner (2016) Exposure to the synthetic progestin, 17α-hydroxyprogesterone caproate during development impairs cognitive flexibility in adulthood. *Endocrinology,* 157**,** 77-82.

Wood, R., D. Clark, A. King, D. Mackay & J. Pell (2013) Novel cross-sectoral linkage of routine health and education data at an all-Scotland level: a feasibility study. *The Lancet,* 382**,** S10-S10.

Wood, S., S. Cooper & S. Ross (2014) Does induction of labour increase the risk of caesarean section? A systematic review and meta-analysis of trials in women with intact membranes. *BJOG: An International Journal of Obstetrics & Gynaecology,* 121**,** 674-685.

World Health Organsiation: Born Too Soon (2012) The Global Action Report on Preterm Birth. Geneva: World Health Organization.

Yang, S., R. W. Platt & M. S. Kramer (2010) Variation in child cognitive ability by week of gestation among healthy term births. *American journal of epidemiology,* 171**,** 399-406.

Yudkin, P., L. Wood & C. Redman (1987) Risk of unexplained stillbirth at different gestational ages. *The Lancet,* 329**,** 1192-1194.

Zeitlin, J., K. Szamotulska, N. Drewniak, A. Mohangoo, J. Chalmers, L. Sakkeus, L. Irgens, M. Gatt, M. Gissler & B. Blondel (2013) Preterm birth time trends in Europe: a study of 19 countries. *BJOG: An International Journal of Obstetrics & Gynaecology,* 120**,** 1356-1365.